# Statistical Modeling of Aircraft Engine Fuel Burn

by

Yashovardhan Sushil Chati

Bachelor of Technology, Indian Institute of Technology Bombay (2012)
Master of Technology, Indian Institute of Technology Bombay (2012)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018



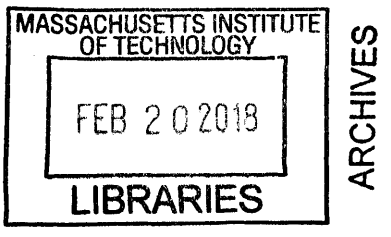
Signature redacted

Author .....
Department of Aeronautics and Astronautics
15 January 2018

Signature redacted

Certified by .....
Hamsa Balakrishnan
Associate Professor, Aeronautics and Astronautics
Thesis Supervisor

Accepted by .....
Hamsa Balakrishnan
Associate Professor, Aeronautics and Astronautics
Chair, Graduate Program Committee



This doctoral thesis has been examined by a committee comprising the following members:

Prof. Hamsa Balakrishnan . Signature redacted .....
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Prof. Tamara Broderick... Signature redacted .....
Assistant Professor of Electrical Engineering and Computer Science
Doctoral Committee Member

Prof. Edward M. Greitzer..... Signature redacted .....
H. N. Slater Professor of Aeronautics and Astronautics
Doctoral Committee Member

Dr. Jayant Sabnis ..... Signature redacted .....
Senior Lecturer of Aeronautics and Astronautics
Doctoral Committee Member

Signature redacted

Prof. Moe Z. Win .....
Professor of Aeronautics and Astronautics
Doctoral Committee Member

# Statistical Modeling of Aircraft Engine Fuel Burn

by

Yashovardhan Sushil Chati

Submitted to the Department of Aeronautics and Astronautics
on 15 January 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

Fuel burn is a key driver of aircraft performance, and contributes to airline costs and aviation emissions. While the trajectory (ground track) of a flight can be observed using surveillance systems, its fuel consumption is generally not disseminated by the operating airline. Emissions inventories and benefits assessment tools therefore need models that can predict the fuel flow rate profile and fuel burn of a flight, given its trajectory data.

Most existing fuel burn estimation tools rely on an architecture that is centered around the Base of Aircraft Data (BADA), an aircraft performance model developed by EUROCONTROL. Operational data (including trajectory data) are generally processed in order to generate the inputs needed by BADA, which then provides an estimate of the fuel flow rate and fuel burn. Although a versatile tool that covers a large number of aircraft types, BADA makes several assumptions that are not representative of real-world operations. Consequently, the reliance on BADA results in errors in the fuel burn estimates. Additionally, existing fuel burn modeling tools provide deterministic predictions, thereby not capturing the operational variability seen in practice.

This thesis proposes an alternative model architecture that enables the development of data-driven, statistical models of fuel burn. The parameters of interest are the instantaneous fuel flow rate (that is, the mass of fuel consumed per unit time) and the fuel burn (cumulative mass of fuel consumed over a particular phase or the entire trajectory). The new model architecture uses supervised learning algorithms to directly map aircraft trajectory variables to the fuel flow rate, and subsequently, fuel burn. The models are trained and validated using operational data from flight recorders, and therefore reflect real-world operations.

A physical understanding of aircraft and engine performance is leveraged for feature selection. An important characteristic of statistical methods is that they provide both estimates of mean values, as well as predictive distributions reflecting the variability and uncertainty. Locally expert models are developed for each aircraft type and for each of the flight phases. The Bayesian technique of Gaussian Process Regression (GPR) is found to be well-suited for modeling fuel burn. The resulting models are found to be significantly better than state-of-the-art aircraft performance models in predicting the fuel flow rate and fuel burn of a trajectory, giving up to a 63% improvement in total airborne fuel burn prediction over the BADA model.

Finally, the Takeoff Weight (TOW) of an aircraft is recognized as an important variable for determining the fuel burn. The thesis therefore develops and evaluates a methodology to estimate the TOW of a flight, using trajectory data from its takeoff ground roll. The proposed statistical models are found to result in up to a 76% smaller error than the Aircraft Noise and Performance (ANP) database, which is used currently for TOW estimation.

Thesis Supervisor: Hamsa Balakrishnan
Title: Associate Professor, Aeronautics and Astronautics

# Acknowledgments

I dedicate my thesis to *maharSi pANini*, the (around) $6^{th}$ century BCE Indian grammarian, who codified the entire Sanskrit Grammar in his revolutionary text *aSTAdhyAyI*. His ancient work of recognizing patterns in the ocean of Sanskrit language and distilling them into less than 4,000 aphorisms is simply incredible. I was fortunate to learn about *aSTAdhyAyI* during my time at MIT. As this thesis has used machine learning techniques to understand patterns in aircraft engine fuel burn from flight data, my thesis dedication to this ancient language scientist is befitting.

It is said that there is nothing as pure as knowledge in this world. My doctoral research has indeed been a journey of learning, of venturing into uncharted waters with the aim of expanding the horizons of knowledge. Many people have been instrumental in supporting me in this journey of over 5 years and as I near my destination, I would like to acknowledge their help and support.

I would, first and foremost, like to thank my advisor, Prof. Hamsa Balakrishnan without whose support, this thesis would not have seen the light of the day. She has been a very encouraging mentor over the past many years and discussions with her opened new doors whenever I got stuck. She has been enormously responsible for my growth as a researcher. I am also thankful to the other members of my doctoral committee, Prof. Edward Greitzer, Prof. Moe Win, Prof. Tamara Broderick, and Dr. Jayant Sabnis for giving new ideas to work on and for teaching me to not lose sight of the big picture while being engrossed in the technical aspects of research. I also thank Dr. Tom Reynolds and Prof. Saurabh Amin for graciously accepting to be my thesis readers.

I am extremely grateful to all the agencies who funded the research and my stay at MIT. This research has been supported in part by the National Science Foundation (NSF) and in part by the Federal Aviation Administration (FAA). I would also like to thank the Transportation Research Board (TRB) for awarding me a scholarship through the 2016-2017 Graduate Research Award Program on Public-Sector Aviation Issues. I am thankful to Monica Alcabin, Bob Arnott, Larry Goldstein, Brian Kim, and Saadat Syed for having reviewed my research under the auspices of the Graduate Research Award Program and to Mary Sandy and Sarah Pauls for managing the logistics of the program. The anonymous reviewers who reviewed my other conference submissions also deserve my sincere acknowledgments for having given new ideas to proceed forward.

My colleagues in the International Center for Air Transportation have played a huge role in my successful journey towards a PhD. They always ensured that the atmosphere in the office was conducive to research. I am thankful to Harshad, Lishuai, Akshay, Tamas, Sathya, Luke, Mike Wittman, Hector, Jacob, Patrick, Karthik, Joao, and Jackie for the wonderful discussions and fun moments spent together. I am particularly thankful to Sandeep Badrinath for being an amazing collaborator on the FAA project, friend, and source of inspiration.

I am indebted to the staff of the Department of Aeronautics and Astronautics for promptly providing me logistical help whenever needed. Thank you, Beth for being a wonderful graduate program administrator and my go-to person for clarification of rules and procedures. I am thankful to Quentin and Ping for taking care of all my financial matters.

Having a roof over one's head is a fundamental need of every human being. I am therefore

grateful beyond words to the Tang Hall Graduate Residence for being my home for 5 years, and to its heads of house (Dawn and Larry Anderson), house manager (Michael Collins), and staff for providing me with all the amenities necessary for a comfortable living. I also got the golden opportunity to serve on the officer board of Tang Hall and in the process improve my organizational and leadership skills. I am going to terribly miss the exercise room in Tang Hall which has been vital in improving my fitness levels during my stay at MIT.

Life at MIT would have been tough without the constant support of my roommates and friends. I will always cherish the wonderful moments I spent with them. Thought-provoking discussions with them have matured my thinking process, which has positively impacted my research too. I would like to place on record my appreciation for Vaibhav, Sanket, Sameer, Parnika, Naga, Tapovan, Divya, Krithika, Ashwin, Jaichander, Deepak for being sources of inspiration in various ways and for making me a better human being. I am particularly grateful to Apoorvaa, Pritish, and Devendra for motivating me when the chips were down and for being my trusted advice-seekers for any problem.

I was fortunate to get exposure to different extra-curricular activities during my doctoral studies, such as playing badminton and chess, learning Sanskrit, or improving my public speaking skills at the Toastmasters club. I am thankful to all my peers involved in these activities for providing me a refreshing change when research became too stressful. I also got an opportunity to attend the MIT-Imperial College Global Fellows Program aimed at improving communication, presentation, and collaboration skills, for which I am grateful to the Global Education and Career Development center at MIT. Many thanks are also due to the Teaching and Learning Laboratory at MIT for having imparted education about modern teaching techniques through their Kaufman Teaching Certificate Program.

The PhD is the culmination of not only 5 years at MIT but 25 years of formal education. I am indebted to all my teachers who, over my lifetime, have taught me several valuable lessons and moulded me into the person I am today. I would like to thank my parents, Sharawati and Sushil Chati, for raising me with utmost care and affection, many times in unfavorable circumstances, and for making me realize the importance of discipline in life. I am also grateful to my uncle Shreerang, and grandparents Anuradha and Arvind, for being beacons of inspiration all my life. Words cannot describe Krishnaa and Baadal, my four-legged siblings, who have been living embodiments of unconditional love in this increasingly materialistic world.

I would like to end by remembering the Almighty under whose watchful gaze I always live.

# Contents

# List of Figures

# List of Tables

# Nomenclature

*Symbols*:

| | |
|---|---|
| $a$ | aircraft longitudinal acceleration |
| A/C | aircraft |
| $a_{\text{thres.}}$ | acceleration threshold in taxi-out trajectory smoother |
| $C_{D_0}$ | lift-independent aircraft drag coefficient |
| $C_{D_2}$ | lift-induced aircraft drag coefficient |
| $C_{f1}$, $C_{f2}$, $C_{f3}$, $C_{f4}$ | BADA TSFC coefficients |
| $C_L$ | aircraft lift coefficient |
| COV | covariance |
| $C_{Tc,1}$, $C_{Tc,2}$, $C_{Tc,3}$, $C_{Tc,4}$, $C_{Tc,5}$ | BADA climb thrust coefficients |
| $C_{\text{Tdes,high}}$, $C_{\text{Tdes,low}}$, $C_{\text{Tdes,app}}$, $C_{\text{Tdes,ld}}$ | BADA descent thrust coefficients |
| $D$ | aircraft drag |
| $\mathcal{D}$ | dataset |
| $f$ | mode fraction; generic function |
| $F_n$ | aircraft net thrust |
| $F_{00}$ | maximum ISA sea level static engine thrust |
| $F_0$ | static engine thrust |
| $f_r$ | ground frictional force |
| $g$ | acceleration due to gravity (= 9.81 m·s$^{-2}$) |
| $\mathcal{GM}$ | Gaussian Mixture distribution |

| | |
|---|---|
| GP, $\mathcal{GP}$ | Gaussian Process |
| $h$ | aircraft pressure altitude |
| $\dot{h}$ | aircraft vertical speed |
| $\mathcal{H}_0$ | null hypothesis |
| $\mathcal{H}_A$ | alternate hypothesis |
| $h_{\text{de}}$ | BADA transition altitude |
| $\mathbf{I_n}$ | $n \times n$ identity matrix |
| i.i.d. | independent, identically distributed |
| $\mathbf{K}$ | covariance matrix |
| $K_1, K_2, K_3, K_4$ | SFI model coefficients (climb out) |
| $k(\mathbf{x}, \mathbf{x}')$ | kernel/covariance function at inputs $\mathbf{x}$ and $\mathbf{x}'$ |
| $L$ | aircraft lift |
| $\boldsymbol{\ell}$ | vector of length scales |
| $M$ | Mach number |
| $\mathcal{M}$ | number of modes |
| $m$ | aircraft mass |
| $m_e(\mathbf{x})$ | GP mean function at $\mathbf{x}$ |
| $\dot{m}_f$ | average fuel flow rate per engine |
| $m_f$ | mass of fuel consumed (fuel burn) |
| $\dot{m}_{f\,\text{min}}$ | BADA minimum aircraft fuel flow rate from all engines |
| $m_s$ | number of inducing inputs in GPR |
| $N$ | ground normal reaction |
| $\mathcal{N}$ | normal/Gaussian distribution |
| $n$ | number of observations |
| $N_{\text{eng}}$ | number of engines on the aircraft |
| $P$ | pressure |
| $p$ | probability distribution function; number of input features |
| $q$ | aircraft dynamic pressure |
| $R$ | gas constant for air (= 287.05 J/kg/K) |

| | |
|---|---|
| $\mathbf{R_n}$ | covariance of additive noise |
| $S$ | distance covered during takeoff ground roll |
| $\mathcal{S}$ | reference wing area |
| $s$ | estimated standard deviation |
| $T$ | temperature |
| $\mathcal{T}$ | Student's t-distribution |
| $t$ | time |
| $V$ | aircraft true airspeed |
| $V_{\mathrm{GS}}$ | aircraft ground speed |
| $V_{\mathrm{GS}_1}$ | aircraft ground speed at start of takeoff ground roll |
| $V_{\mathrm{GS}_2}$ | aircraft ground speed at end of takeoff ground roll |
| $\mathbf{v_n}$ | additive noise vector |
| $\mathbf{W_n}$ | covariance of process noise |
| $\mathbf{w}$ | vector of weights (statistical sense) |
| $w_a$ | component of process noise for acceleration |
| $\mathbf{w_n}$ | process noise vector |
| $w_\theta$ | component of process noise for turn rate |
| $\mathbf{X}$ | matrix of input vectors |
| $\mathbf{x}$ | input vector; state vector |
| $(x, y)$ | aircraft coordinates on the ground |
| $y$ | output |
| $\mathbf{z}$ | measurement vector |
| 0.95PI | 95% prediction interval |
| $\alpha, \beta_1, \beta_2, \beta_3$ | SFI model coefficients (approach) |
| $\alpha_s$ | level of statistical significance |
| $\beta$ | OLS regression parameters |
| $\beta_{\mathrm{ISA}}$ | ISA tropospheric temperature lapse rate (= -0.0065 K/m) |
| $\gamma$ | adiabatic constant for air (= 1.4) |
| $\gamma_c$ | flight path angle |

| | |
|---|---|
| $\Delta$ | change |
| $\delta$ | corrected pressure (corrected by the ISA sea level static pressure = 101,325 Pa); Kronecker delta |
| $\epsilon$ | noise |
| $\zeta$ | degrees of freedom of the Student's t-distribution |
| $\eta$ | thrust deration level |
| $\theta$ | corrected temperature (corrected by the ISA sea level static temperature = 288.15 K) |
| $\theta_c$ | aircraft course track angle on the ground |
| $\theta_h$ | GPR hyperparameter vector |
| $\mu$ | mean of distribution |
| $\mu_r$ | coefficient of friction |
| $\nu$ | efficiency |
| $\rho$ | air density |
| $\sigma$ | standard deviation of distribution |
| $\omega$ | aircraft turn rate |

*Subscripts*:

| | |
|---|---|
| ANP | using the ANP model |
| (ap) | in approach |
| (as) | in full ascent |
| ATD | above touchdown |
| BADA | using the BADA model |
| (co) | in climb out |
| (cr) | in cruise |
| (de) | in full descent |
| DP | using the dot product kernel |
| DPE | using the dot product exponential kernel |
| DPM, 3/2 | using the dot product Matérn kernel with parameter 3/2 |

| | |
|---|---|
| DPM, 5/2 | using the dot product Matérn kernel with parameter 5/2 |
| DPSE | using the dot product squared exponential kernel |
| eng | engine |
| GPR | using the GPR model |
| $i$ | $i^{th}$ index (instant, observation) |
| ICAO | using the ICAO aircraft engine emissions databank |
| ICAO – BFFM2 | using the ICAO databank with BFFM2 corrections |
| Mat., 3/2 | using the Matérn kernel with parameter 3/2 |
| Mat., 5/2 | using the Matérn kernel with parameter 5/2 |
| pe | per engine |
| pred | prediction |
| $s$ | samples |
| SE | using the squared exponential kernel |
| SFI | using the SFI model |
| SL | at sea level |
| surr | surrogate |
| taxi | in taxi |
| TO | at takeoff |
| tot | total |
| true | ground truth value |
| $\infty$ | ambient air |
| – (minus) | excluded variable |

*Superscripts*:

| | |
|---|---|
| T | transpose |
| * | in prediction set (not seen in training) |
| $\hat{}$ | estimate |
| $\dot{}$ | time derivative |

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Chapter 1

# Introduction

Since the Wright brothers successfully demonstrated controlled flight by a heavier-than-air aircraft for the first time in December 1903, air traffic has evolved into a system with 1,400 airlines, operating 32.8 million commercial flights annually, from nearly 3,900 airports worldwide, and using more than 26,000 aircraft [2]. Aircraft transported 3.6 billion passengers and 51.2 million tonnes of freight in 2015. Global passenger traffic has seen annual growth rates of 7.3% (in terms of revenue passenger-miles) in the last few years, while passenger enplanements are forecast to grow by 2.3% per year over the next two decades [3, 4].

Fuel burn constitutes a major component of the Direct Operating Costs of an airline, and is an important aspect of aircraft and engine performance. Airline operations consumed 81 billion gallons of fuel in 2015, which corresponded to 26.5% of total airline expenses [2]. Fuel burn also generates emissions of pollutants that have adverse impacts on climate, air quality, and health. Aircraft are estimated to add about 773 million tonnes of carbon dioxide into the atmosphere every year [2]. Increased air traffic demand in the United States is projected to result in a 2–3.6 fold increase in $CO_2$ emissions, and a 1.2–2.7 fold increase in nitrogen oxide emissions, between 2000 and 2050 [5]. Although aircraft contribute to just 3.4% of the total manmade $CO_2$ emissions in the US, these emissions have amplified impacts due to the high altitudes at which they occur [4].

While fuel burn is widely accepted to be an important metric of system performance, the actual per-flight fuel consumed by particular flights between two airports, or flying specific

procedures, is not disseminated by airlines. There are a variety of reasons for this, ranging from concerns about privacy, to competitive interests. The rate at which fuel is consumed by an aircraft is typically archived in the onboard Flight Data Recorder (FDR), along with a range of other flight parameters. Many airlines have *internal* Flight Operations Quality Assurance (FOQA) programs in which they routinely analyze flight data, in order to improve safety and efficiency, and reduce maintenance costs [6]. However, there are significant restrictions against broadly sharing these data, amid concerns that they could be potentially used for disciplinary actions against pilots [7]. There are also sensitivities surrounding other flight parameters that influence fuel burn (for example, the weight of the aircraft), which can reveal airline business strategies such as load factors and fuel tankering practices. In an industry with low margins (the net profit margin of the airline industry in 2016 was a historically high 5.1%; the average over the previous 5 years was only 2% [8]), airlines are hesitant to share information that could reveal any competitive advantage. As a result of all these factors, there is a need for modeling tools that can estimate the fuel burn corresponding to real-world air traffic operations.

## 1.1 Motivation: Existing models of fuel burn

The estimation of the fuel burn impacts of operations has been a long-standing challenge in Air Traffic Management (ATM) [9]. The motivation behind this challenge is two-fold, namely, to evaluate the impact of real-world operations (for example, to determine inventories of fuel burn and emissions [10]), as well as to evaluate the benefits of operational changes and system modernization efforts (for example, NextGen in the US [11], SESAR in Europe [12], and the Australian Air Traffic Management Plan [13]). The need for fuel burn estimation has led to the development of several fuel burn models, including the Federal Aviation Administration (FAA)'s *NextGen Office Model* [1], *Aviation Environmental Design Tool (AEDT) widely used in the US to model aircraft fuel burn, emissions, and noise* [14], and *Aircraft Fuel Evaluation Simulation Tool (AFEST)* [15], the *EJPM-based Trajectory Analysis Software (ETAS)* developed by GfL and Technische Universität Dresden [16], and operationally used by the German Air Navigation Service Provider (ANSP) Deutsche Flugicherung (DFS), and Airservices Australia's

*Dalí* [17].

### 1.1.1 Model architectures

While these fuel burn models have generally been developed in-house by different parts of the ATM community, there has been a remarkable similarity in the resulting architectures [1]. Figure 1-1, adapted from recent work by Enea et al. [1], presents a flowchart of the typical architecture of these fuel burn models. The main inputs, shown as parallelograms, are *surveillance data* or the flight tracks to be evaluated, *user preferences* which reflect airline and air traffic operational behavior, and the *Takeoff Weight (TOW)* of each flight. The ultimate desired output is the *fuel burn* (mass of the fuel consumed over time) estimate for each of the flights, while the intermediate output is the *fuel flow rate* (mass of fuel consumed per unit time) of each flight at any point in time. These outputs are depicted in blue. Perhaps the most striking common feature of these fuel burn models, as well as several others not mentioned above [18, 19, 20], is their use of EUROCONTROL's Base of Aircraft Data (BADA) [21] as an underlying Aircraft Performance Model (APM). The corresponding portion of the flowchart is highlighted in orange in Figure 1-1.

#### Base of Aircraft Data (BADA)

BADA is an APM that has been developed and refined, over the past two decades, for the "simulation and prediction of aircraft trajectories for purposes of ATM research and operations" [21, 22, 23]. It specifies, for each aircraft type in the database, performance and operating procedure coefficients that can be used to calculate aerodynamic and engine characteristics such as drag and thrust, and subsequently the fuel flow rate. The latest release in BADA Family 3, namely, BADA 3.13, includes aircraft models for 519 different aircraft types. BADA 4.1, which has considerable modeling refinements, covers 73 aircraft types, but at a higher level of precision than BADA Family 3 [21]. The BADA models for a particular aircraft type are essentially obtained by fitting polynomials to operational data obtained from aircraft manufacturers using least-squares techniques [24, 25]. These models (coefficients) are provided for each aircraft type, for a discrete set of aerodynamic configurations (for example, takeoff, climb, cruise,

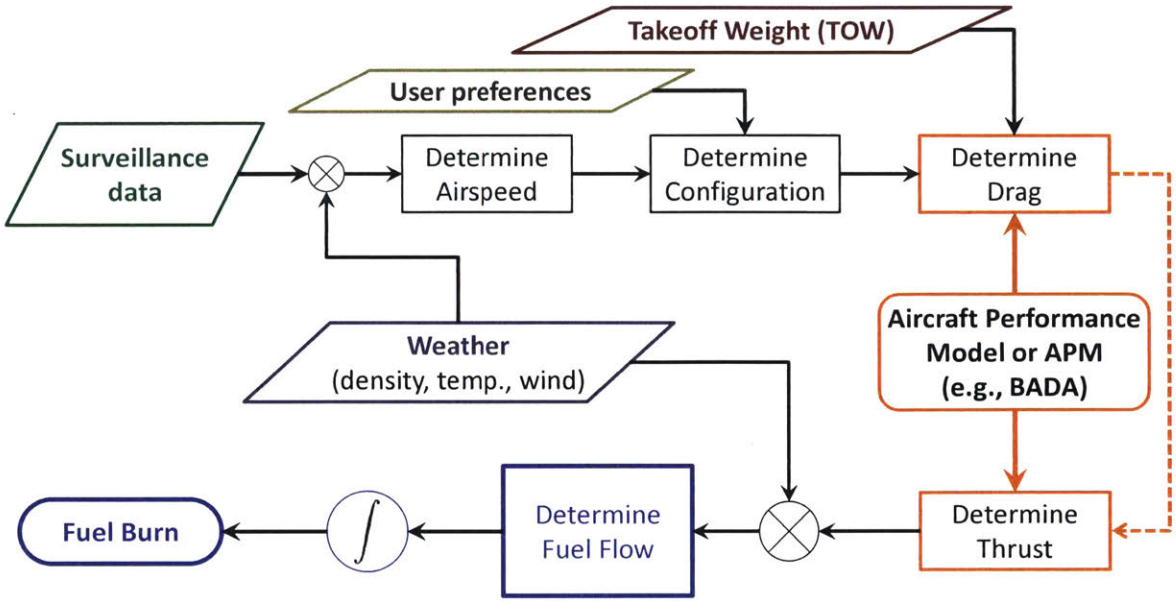



Figure 1-1: Flowchart illustrating typically-used architectures for determining fuel burn from trajectory surveillance data (adapted from [1]).

descent, approach, and landing), and a set of candidate mass levels (low, nominal, high). The configuration can be driven either by default values, or by *user preferences*, which can be used to reflect airline procedures or air traffic control procedures. Finally, we note that BADA assumes a knowledge of the mass of the aircraft, which in turn requires an estimate of the mass of the aircraft at takeoff (known operationally as the Takeoff Weight, or TOW).

### 1.1.2 Limitations of current state-of-the-art

Despite their widespread use, these existing fuel burn models pose significant limitations. BADA was initially developed as a trajectory prediction and simulation tool in the 1990s (BADA 2.1 was released in May 1994 [22]), when aircraft trajectory data were not systematically collected or archived. Consequently, a key purpose of these models was to generate the complete 4-D (space and time) trajectory of a flight between two airports, given the takeoff mass and a nominal speed schedule (which could be used to determine when the aircraft would transition from climb to cruise, cruise to descent, and so on). BADA's Total Energy Model (TEM) then balances the rate of work done by the different forces (thrust and drag) acting on the aircraft to the rate of

increase in the total (potential + kinetic) energy in order to determine approximations of the rate of climb/descent (ROCD), engine thrust, fuel flow rate, etc. The coefficients of polynomial fits are determined with the objective of minimizing the errors in ROCD and fuel flow rate. Since the drag forces depend significantly on the configuration of the aircraft (for example, due to the different flap settings during approach and landing), separate model coefficients are specified for different phases of flight [22].

In general, existing models for fuel burn assume a knowledge of parameters which are difficult to obtain or estimate (such as flap schedules and the drag profiles). They also employ simplified parametric equations which neither adequately capture the physical underpinnings nor the complexity in system behavior (for example, BADA Family 3 models the Thrust Specific Fuel Consumption (TSFC) (ratio of the fuel flow rate to the net thrust) as a linear function of the airspeed, neglecting its dependence on ambient conditions and net thrust). The underlying data are derived from non-operational sources such as flight manuals, ground tests, handbooks, performance calculators. As a result, such studies do not reflect the effects of operational variability, random disturbances affecting aircraft performance, and other unmodeled factors. There are additional limitations with regards to their ability to incorporate trajectory as well as operational data, as discussed in the next few sections.

### Operational variability

Aircraft operations occur in constrained, complex environments, which result in differences between the predictions of idealized models, and the actual fuel burn. The reasons for such operational variability include differences in weather conditions (such as winds), aircraft weights, fueling practices, maintenance operations, age of the aircraft, air traffic control procedures, etc. As a consequence of these factors, two aircraft of the same type, flying very similar trajectories, could consume significantly different amounts of fuel. Airlines have reduced the Zero Fuel Weight (ZFW) of their aircraft by moving to paperless cockpits, reducing excess potable water, reducing galley equipment, and even removing seatback phones and magazines: It is estimated that a 1,000 lb reduction in the ZFW can result in a fuel savings of 0.6–0.7% for a Boeing 737 [26, 27]. Similarly, the Operating Empty Weight (OEW) of an aircraft is estimated to increase

0.1–0.2% per year, resulting in aircraft being as much as 1% heavier when they are 5–10 years old, due to the accumulation of moisture and dirt. Maintenance activities, such as repetitive engine washes, can also have a significant impact on fuel consumption [27]. Given an aircraft trajectory, we would like to predict not just the expected value, but also the variability in the fuel burn that would be seen in actual operations.

### Incorporation of aircraft trajectory data

In the years since BADA was first developed, aircraft trajectory data have become widely available. Trajectory data can be gathered by radar surveillance (for example, the FAA's Enhanced Traffic Management System (ETMS), Aircraft Situation Display to Industry (ASDI) (both now a part of the Traffic Flow Management System (TFMS)), Performance Data Analysis and Reporting System (PDARS), and Airport Surface Detection Equipment-Model X (ASDE-X), EUROCONTROL's Enhanced Tactical Flow Management System (ETFMS)), or from Automatic Dependent Surveillance-Broadcast (ADS-B) ground receivers. (These systems are the data-sources for many of the commercial real-time flight-tracking products that are available today.) The availability of surveillance data has resulted in the need to estimate the fuel burn impacts of the trajectories, as they were flown by actual aircraft [1]. However, even with the incorporation of surveillance data, the basic architecture for fuel burn estimation has not changed. Model refinements have primarily focused on improving different parameters within the existing architecture.

### Refinement of fuel burn models using operational data

The factors described in Section 1.1.2 result in significant differences between the fuel burn estimated by models and that observed in actual operations. This fact has motivated the use of operational data to refine models of fuel burn. However, prior efforts have focused on either evaluating the outputs or tuning various parameters of the existing architecture (Figure 1-1), rather than on investigating new ones. Data from FDRs as well as aircraft manufacturer manuals and handbooks have been used to refine coefficients in BADA models [24], develop speed and flap deployment schedules [16, 17], tune the thrust determination models of the APM [28],

and improve TOW estimates [17, 29]. These efforts have also focused on deterministic (point) estimates of the fuel flow rate and fuel burn. By contrast, this thesis investigates an alternative architecture in which the fuel flow rate and fuel burn may be directly inferred from trajectory data, along with estimates of their variability.

## 1.2 Main contributions of this thesis

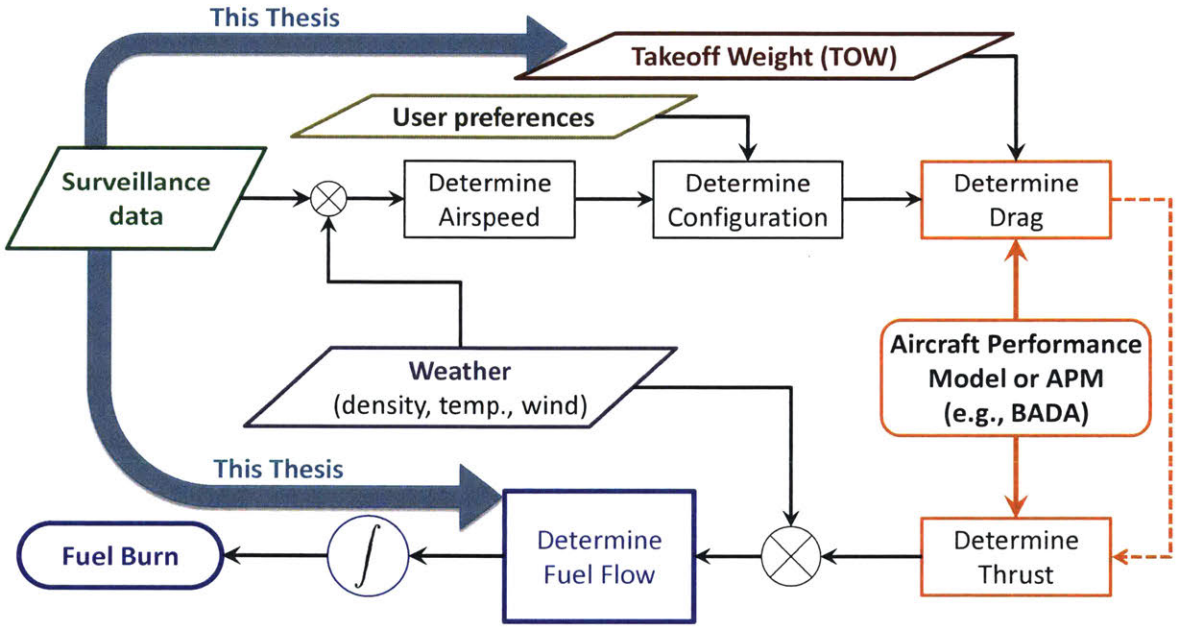



Figure 1-2: Flowchart of typically-used architectures for determining fuel burn from trajectory surveillance data (adapted from [1]), as well as the new approach pursued in this thesis (depicted by the teal block arrows).

The overarching research contribution of this thesis is the development of methodologies for the statistical modeling of aircraft engine performance. It focuses on two specific aspects of engine performance, namely, fuel flow rate and fuel burn. At a high-level, this thesis investigates the following problem: *Given the trajectory of a flight, how does one infer its fuel flow rate and fuel burn profiles, along with their associated uncertainties?*

The main parameters of aircraft engine performance investigated in this thesis are the instantaneous fuel flow rate (mass of fuel consumed per unit time) and the fuel burn (mass of the

fuel consumed over time). Both these quantities are important from an emissions perspective: Some emissions (e.g., nitrogen oxides, hydrocarbons) depend on the instantaneous fuel flow rate, whereas others (e.g., carbon dioxide, water vapor) depend on the total mass of fuel burnt. The fuel flow rate and the fuel burn are modeled in all the phases of flight (sometimes referred to as a *gate to gate* [30]), both on the ground and in the air. This is in contrast to many existing tools of fuel burn estimation which focus only on the airborne phases of flight.

### Statistical models of fuel burn

The models developed in this thesis estimate the fuel flow rate profiles given the trajectory information of a flight. The underlying mappings depend on a multitude of components, such as the aircraft and engine types, weather conditions, operational procedures, human factors, etc. Statistical methods overcome the challenges described in Section 1.1.2 by providing predictive distributions of the fuel burn of a particular trajectory. In short, they yield mean estimates as well as prediction intervals for the fuel burn as part of the model output. The resulting prediction intervals reflect the cumulative effect of random disturbances affecting engine operation (such as manufacturing tolerances, component deterioration, and atmospheric and flow turbulence [31]), as well as features neglected in model building. The model predictions are therefore representative of both the variability in fuel burn and model uncertainty.

We propose methods to infer the *direct* mapping between the trajectory and the corresponding fuel flow using data from real operations, in order to address the challenges described in Section 1.1.2. We adopt a *supervised learning* approach, namely, given training datasets of trajectories and their fuel flow, we build models that can be used to estimate the fuel flow profiles of new trajectories that we are interested in evaluating [32]. As mentioned earlier, FDR archives contain both trajectory and fuel flow variables from real flight operations, making them an ideal source of training data. The developed models can then, given a new set of trajectory variables from a flight, predict the associated fuel burn even in the absence of the FDR data for that flight. In short, the objective of this thesis is to develop *open models*, namely, ones that can estimate the fuel burn of a flight from easily accessible data such as its trajectory variables, thereby addressing the challenges described in Section 1.1.2.

In this thesis, we investigate several classes of machine learning algorithms, including least squares regression, Classification and Regression Trees (CART), Least Squares Boosting (LSB) using regression trees, and Gaussian Process Regression (GPR). We find that the Bayesian technique of GPR presents several advantages in the problems of interest. Locally expert models are developed for each aircraft/engine type and for each of the flight phases. While the models are not explicitly physics-based, we use a physical understanding of aircraft and engine performance in order to identify the predictor variables or features. In order to enable practical implementation, the features are restricted, to the extent possible, to those that are derivable from trajectory data.

The data-driven statistical models are compared for their predictive performance with other widely used Aircraft Performance Models (APMs) and are found to exceed them in predicting the fuel flow rate and fuel burn, as well as their uncertainties, for a new flight trajectory. In particular, the GPR-based models are shown to yield a median (across all aircraft types) value of the mean absolute fuel burn error of 2.2% in ascent, 2.0% in climb out, 7.2% in cruise, 8.2% in descent, and 5.5% in approach. The GPR models are shown to achieve a reduction in mean absolute fuel burn error of as much as 54% in ascent, 80% in climb out, 34% in cruise, 66% in descent, and 89% in approach as compared to currently used APMs.

The total fuel consumed by a trajectory can be predicted by aggregating the fuel flow rate predicted at each time instant. By doing so, we find that the proposed GPR-based models achieve a mean error of 0.2% (across flights of 8 different aircraft types), and a mean absolute error of 3.8% in the total airborne fuel burn. These results constitute a significant improvement over existing fuel burn models (for example, an up to 63% improvement over the BADA model).

Finally, model performance on flight surveillance data from the airport surface is described, in order to demonstrate the practical applicability of the developed models. The modeling approaches adopted in this thesis are descriptive ones, that is, they use data to identify mappings between the input and output variables. They do not attempt to identify causal relationships between the input and output variables, nor prescribe optimal operational procedures. As new operational data become available, the same methodologies can easily be applied to refine models of fuel burn estimation, or even develop models of other aircraft performance parameters of

interest.

**Modeling of takeoff weight**

Consistent with prior research, the aircraft Takeoff Weight (TOW) is found to be an important, yet generally not available, variable in determining the fuel burn [1]. TOW is also important for aircraft trajectory modeling and estimation of other parameters such as aircraft climb performance [33]. This thesis therefore, also develops a methodology to estimate the takeoff weight directly using trajectory information from the takeoff roll. In particular, it proposes a statistical approach based on Gaussian Process Regression (GPR) to determine both a mean estimate of the TOW and the associated prediction interval, using observed data from the takeoff ground roll. Following a similar philosophy to the fuel burn modeling, predictor variables are chosen by considering both their ease of availability and the underlying aircraft dynamics, with the FDR data providing the ground truth for a supervised learning approach. The proposed models are found to have a mean absolute error in TOW prediction of 3.6% (averaged across nine different aircraft types), resulting in a nearly 35% smaller error than the models in the Aircraft Noise and Performance (ANP) database that is used in AEDT [14, 34]. In contrast to the ANP database which provides only point estimates of the TOW, the GPR models quantify the uncertainty in the estimates by providing a predictive distribution.

### 1.2.1 Intended benefits

Fuel burn is an important aspect of operational performance, and is therefore of interest to many stakeholders in the ATM community. A key challenge lies in evaluating the relative fuel burn impacts of different procedures and routes. The ideas developed in this thesis represent the use of state-of-the-art machine learning techniques in order to better predict a flight's fuel burn from its trajectory, thereby enhancing current capabilities for modeling aircraft and engine performance. The models developed in this thesis are computationally less expensive than full-blown physical simulations of aircraft performance, and can yield predictive distributions without the need for identification of uncertain model parameters and subsequent Monte Carlo simulations. The proposed models can enable airline flight planners to assess the fuel burn

and emissions impact of different routes. In addition, these methodologies can be extended in the future to address the problem of anomaly detection for aircraft maintenance operations, that is, by identifying flights whose fuel burn profiles deviate significantly from the nominal values. In the near-term, these models will help improve fuel burn and emission inventories by evaluating impacts on a per-flight operation basis (instead of on a fleet/airline/regional level alone). Finally, these models have the potential to help analyze the effects of new operational procedures to mitigate the environmental impacts of aviation.

## 1.3 Thesis organization

The structure of the remainder of this thesis is as follows: Chapters 2–4 complete the preliminaries, with Chapter 2 presenting prior research and the state-of-the-art in fuel burn estimation, Chapter 3 providing details of the different datasets used for model development and evaluation, and Chapter 4 describing the fundamentals of different statistical algorithms investigated in this thesis. The next four chapters constitute the main portion of the thesis. Chapter 5 focuses on modeling the fuel flow rate and mass of fuel consumed in the airborne phases of flight, including the process of feature identification, model training, validation, and testing, and comparisons to other APMs. The work on fuel burn modeling in the climb out and approach phases was conducted as part of an Airport Cooperative Research Program (ACRP) Graduate Research Award on Public Sector Aviation Issues, and will appear in [35]. Preliminary results on fuel flow rate estimation (in particular, the one-step prediction algorithm described in Section 5.3.2) appeared in [36]. Chapter 6 deals with the modeling of fuel burn on the airport surface during taxi. It also demonstrates the application of the models to surveillance data from ASDE-X. Chapter 7 addresses the estimation of an aircraft's Takeoff Weight (TOW) from takeoff roll data. A version of this work was published in [37]. Finally, Chapter 8 summarizes the key findings of the thesis, and also discusses limitations of this work and promising directions for future research.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Chapter 2

# Related Literature

This chapter describes prior literature related to the modeling of aircraft fuel burn, with the objective of identifying the state-of-the-art, as well as gaps in knowledge that could be addressed in this dissertation.

## 2.1 Simulation models

Simulation models (also known as knowledge-driven, physics-based, behavioral, or process models) are often used to analyze system performance. These models use a physical understanding of system performance to derive modeling equations. Such models can pose challenges when applied to complex systems such as aircraft.

Engine performance simulators (e.g., GasTurb) apply principles of aerodynamics and thermodynamics to engine flows to determine the engine performance metrics (such as fuel flow rate, thrust) at different operating conditions [38, 39]. These simulators require as inputs parameters such as the engine operating point (determined by the ratio of the total temperatures at the high-pressure turbine inlet and the low-pressure compressor inlet), the amount of bleed air, and the power offtake from the compressor. As a result, these models are the most effective in design studies, where the underlying simulation variables (inputs) can be controlled. They are less suitable for estimating the fuel burn of real flight operations, since the simulations would in turn require that these inputs be inferred. Moreover, engine simulations need access

to component characteristics maps, which are proprietary to engine manufacturers.

## 2.2 Data-driven models

Most modern engineering systems are instrumented with embedded sensors, controllers, actuators, and computational components. Such systems are also known as Cyber-Physical Systems (CPS) [40, 41]. The sensors in a CPS provide data which can give valuable insights into system performance. When combined with improvements in storage and computational technologies as well as machine learning algorithms, these data sources can be used to develop better models. Such data-driven modeling approaches have shown promise in diverse applications, such as monitoring battery health [42], hydrological applications [43], energy systems [44], vehicular traffic [45], etc. These studies identify important variables to be used as model inputs, and then apply statistical methods to data in order to relate the output parameters of interest to the inputs. Data-driven methods have the potential to give more realistic estimates of system performance than simulation-based models. One reason is that the observed data are objective manifestations of system performance, and do not depend on modeling assumptions. Data-driven models can also analyze system behavior without requiring a complete understanding of the underlying components.

Data-driven approaches often use statistical machine learning techniques for model development. Regression, classification, and clustering algorithms use data in order to learn mappings between the input and output variables, identify patterns, and determine similarities and dissimilarities in the data. The resulting models are capable of predicting the outputs corresponding to a new set of inputs. An additional advantage of using statistical techniques is their ability to model the system stochastically rather than deterministically. Real-world systems are rarely deterministic, and exhibit considerable variability. For example, an aircraft engine is influenced by random disturbances such as flow turbulence, fluctuations in ambient temperature and pressure, component deterioration and aging, manufacturing tolerances, fuel quality, mechanical vibrations, and drag deterioration [31]. Moreover, there will be unmodeled factors that influence system behavior, leading to uncertainty in the output variables. Deriving the uncer-

tainty estimates for a deterministic model requires (potentially a large number of) Monte Carlo simulations, which in turn requires probability distributions of the input parameters. Such simulations become computationally intractable when there are many uncertain parameters. By contrast, statistical models provide a natural way (by means of prediction intervals) to capture uncertainty and variability in the output variables.

An aircraft (and its engine) can be considered an example of a CPS. There are a number of data sources relating to aircraft performance, including the results of ground tests, manufacturer manuals, and flight recorder datasets. Prior work on data-driven models of aircraft performance has been quite limited, but has included the development of neural network models from aircraft flight performance manuals [46], the application of total energy balance to an aircraft to calculate fuel consumption using data from performance manuals, noise measurement reports, flight operation manuals [9], and the use of nonlinear regression on data from performance calculators in order to model the fuel burn in the terminal phases of flight [28, 47].

## 2.3 Models of aircraft fuel burn

Financial and environmental reasons, as well as the need to assess the fuel burn impacts of various flight operations, have led to aircraft fuel burn estimation being a widely studied problem. In this section, some of the most commonly-used models of fuel burn are discussed. These models form components of broader fuel burn and emissions modeling tools.

### 2.3.1 ICAO Aircraft Engine Emissions Databank

The International Civil Aviation Organization (ICAO) maintains a databank known as the Aircraft Engine Emissions Databank [48] (hereafter referred to as the ICAO Databank). It provides point estimates of the values of fuel flow rates, emission indices (mass of emissions generated per unit mass of fuel burnt), smoke numbers, and times-in-mode at four different thrust settings. These thrust settings correspond to the standard takeoff roll (assumed to last 42 s at 100% thrust), climb out up to 3,000' Above Ground Level (AGL) (132 s at 85% thrust), approach from 3,000' AGL (240 s at 30% thrust), and ground roll and taxi (1560 s at 7% thrust) phases of the

Landing and Take Off (LTO) Cycle. The values are obtained by conducting ground-based tests on an uninstalled engine during the certification process and are corrected to the International Standard Atmosphere (ISA) sea level, static conditions. The values of the fuel flow rates and emission indices are fixed for a particular engine type, irrespective of actual operation. The fuel flow rate values tabulated in the ICAO Databank can be corrected for non-reference atmospheric conditions and installation effects using the Boeing Fuel Flow Method 2 (BFFM2) [10]. The ICAO Databank along with the BFFM2 corrections is used in FAA's Aviation Environmental Design Tool (AEDT) to estimate fuel flow rates in the LTO cycle, when thrust values are not explicitly available.

Studies have shown that the fuel burn estimates from the ICAO Databank method can be significantly different from those derived from operational flight data [49, 50]. The values of the emissions indices reported in the ICAO Databank have also been shown to deviate from measured values [51, 52]. In [49], we compared the times-in-mode, fuel flow rates, and fuel burn values derived from operational flight recorder data in the different phases of the LTO cycle with the corresponding ICAO Databank values. All reported values were converted to sea level static ISA conditions for an uninstalled engine in order to conduct a valid comparison with the ICAO Databank estimates. In most cases, the mean operational values were found to differ statistically significantly (at a 5% significance level using the two-sided Wilcoxon signed-rank test) from those in the ICAO Databank (Figure 2-1). The ICAO Databank was found to overestimate the values of the LTO cycle fuel burn by as much as 39%. These discrepancies were largely due to the underlying assumptions made by the ICAO Databank. For example, the ICAO Databank assumes that takeoff occurs at 100% thrust, whereas in reality, most takeoffs occur at derated thrust levels. Many taxi operations also take place at thrust settings lower than 7% [51]. The differences between the ICAO Databank assumptions and the operational values can lead to inaccuracies in fuel burn and emissions inventories, which currently rely on the ICAO Databank. Moreover, because of its deterministic nature, the ICAO Databank cannot provide uncertainty estimates to reflect real-world operational variabilities.

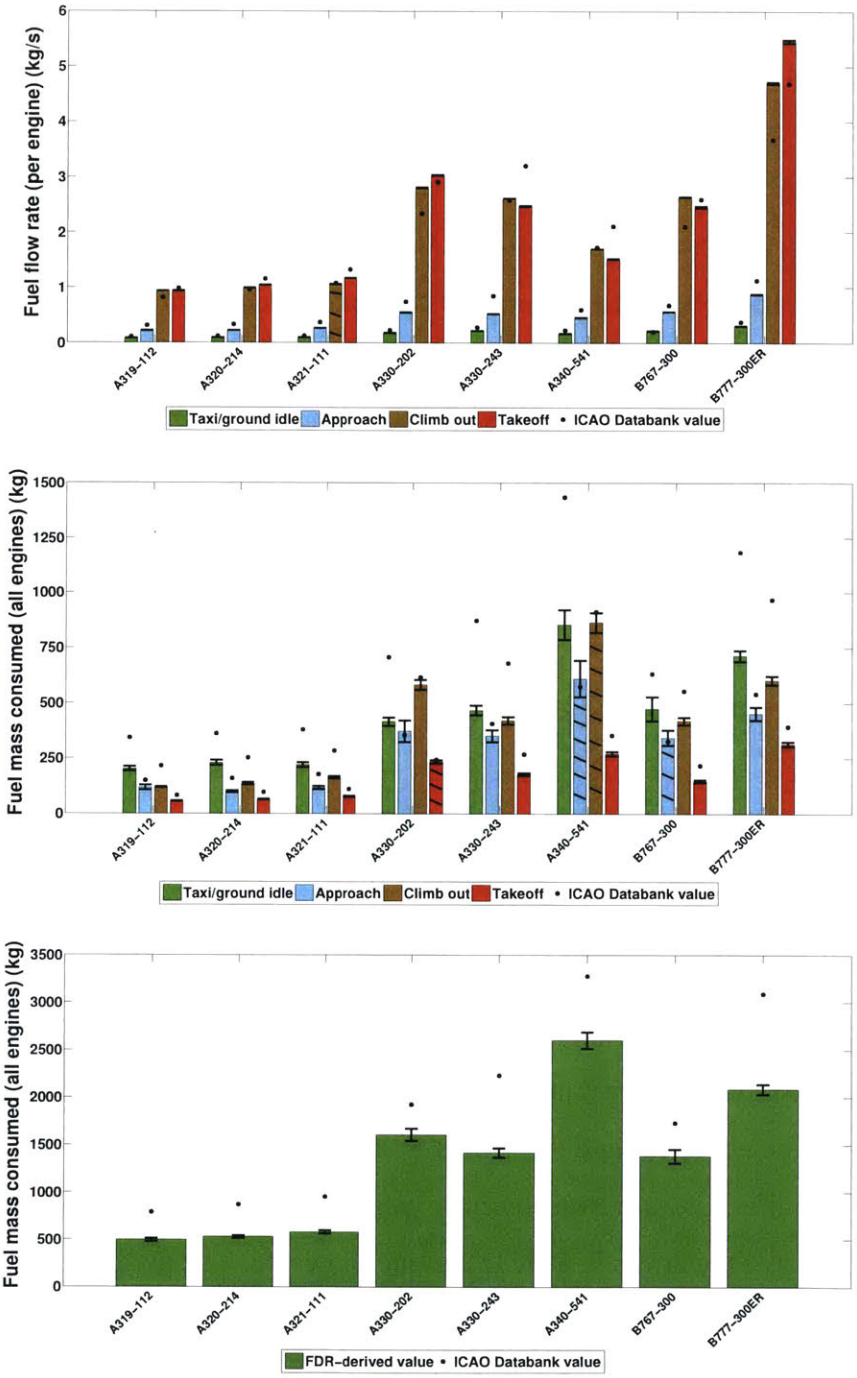



Figure 2-1: LTO cycle: Operational and ICAO Databank values for (top) fuel flow rate (per engine), (middle) fuel mass consumed (all engines), and (bottom) total fuel mass consumed (all engines) in the complete LTO cycle. The bars represent the mean estimates from operational FDR data. The error bars are 95% confidence intervals. Hatched bars represent cases when the operational estimates are *not* statistically significantly different from the ICAO Databank values at a 5% significance level (adapted from [49]).

### 2.3.2 Base of Aircraft Data

Another model commonly used to simulate aircraft performance is the Base of Aircraft Data (BADA), developed by EUROCONTROL using data from different airplane manufacturers and airlines [22]. BADA was initially developed as a trajectory prediction and simulation tool. Consequently, a key purpose of these models was to generate the complete 4-D (space and time) trajectory of a flight between two airports, given the takeoff mass and a nominal speed schedule (which could be used to determine when the aircraft would transition from climb to cruise, cruise to descent, and so on). BADA uses the 'total energy' method to simulate aircraft performance. The Total Energy Model (TEM) balances the rate of work done by the different forces (thrust and drag) acting on the aircraft to the rate of increase in the total (potential + kinetic) energy in order to determine approximations of the rate of climb/descent (ROCD), engine thrust, fuel flow rate, etc. The aircraft is assumed to be a point mass. The governing equation of the TEM is as follows [22]:

$$(F_{n_{\parallel}} - D) \cdot V = mg\frac{dh}{dt} + mV\frac{dV}{dt} \tag{2.1}$$

Here, $F_{n_{\parallel}}$ is the aircraft net thrust from all engines parallel to the aircraft true airspeed vector, $D$ is the drag, $V$ is the true airspeed, $m$ is the aircraft mass, $g$ is the acceleration due to gravity, $h$ is the altitude, and $t$ is the time.

The TEM is used to develop parametric (polynomial) equations to calculate performance parameters such as the net thrust and the Thrust Specific Fuel Consumption (TSFC). The coefficients of polynomial fits are determined empirically using a least-squares technique, with the objective of minimizing the errors in ROCD and fuel flow rate. The performance and operating procedure coefficients used in these equations differ for each aircraft type, and are maintained in a database. BADA Family 3 version 3.13 (hereafter referred to simply as BADA) has 519 different aircraft types in its database. BADA determines the fuel flow rate by estimating the TSFC and the net thrust, and then multiplying them together. The TSFC for jet engines is modeled as a linear function of the airspeed. The minimum fuel flow rate for engine operation is modeled as a linear function of the altitude. The net thrust is modeled as a quadratic function of the air-

craft altitude in ascent and descent and by drag balance in cruise. Since the drag forces depend significantly on the configuration of the aircraft (for example, due to the different flap settings during approach and landing and due to landing gear extension), separate model coefficients are specified for different phases of flight. BADA assumes discrete levels of TOW (such as low, nominal, high, minimum, maximum, reference, and maximum payload) for performance analysis. The BADA fuel flow rate module is used as the Aircraft Performance Model (APM) in several fuel burn estimation tools. These include the FAA's AEDT in the airborne phases [10], Aircraft Fuel Evaluation Simulation Tool (AFEST) [15] and the NextGen Office Model [1], a model developed at NASA [18], Airservices Australia's Dalí [17], and TU Dresden's Enhanced Jet Performance Model (EJPM)-based Trajectory Analysis Software (ETAS) [16].

The BADA models suffer from several drawbacks. BADA APMs are built using data from nominal aircraft operations, which may differ a lot from actual operations. The modeling equations are also overly simplistic. For example, the TSFC is modeled as a linear function of the airspeed alone, while in reality, the TSFC depends on other factors such as ambient conditions and the engine operating point. A linear equation is also insufficient for modeling the complex, nonlinear, engine dynamics. BADA gives point estimates of the fuel burn, and does not provide uncertainty estimates which can quantify the operational variability in the fuel flow rate. It also needs the net thrust values to estimate the fuel flow rates. In practice, this thrust estimation requires the estimation of aircraft drag. The thrust estimation equations in BADA are simplistic parametric equations. Finally, BADA uses a set of candidate mass levels (low, nominal, high) to represent the TOW; consequently, the operational TOW for a particular flight may differ significantly from these discrete levels. As a result of these issues, the estimates of fuel burn provided by BADA can be erroneous. Studies have indicated that the BADA method can give errors as high as 22.3% for phases of flight occurring in the terminal area [28]. More recently, a comparison of various fuel burn models has shown that for a common set of 60 flights, the airborne fuel burn errors ranged from -27.3% to 13.3% for AFEST (median value of -13.1%, with no TOW input), from -31.3% to 26.3% for AEDT (with a median value of -8.6%), from -31.2% to 46.9% for ETAS (with a median value of -11.8%), and from -16.9% to 26.5% for Dalí (with a median value of -3.9%). All these models use BADA 3 as their APM. A version of

Dalí using BADA Family 4 resulted in errors between -13.8% and 32.7%, with a median error of -4.6% [1].

### 2.3.3 Senzig-Fleming-Iovinelli model

Researchers have also developed data-driven models to estimate the TSFC in terminal areas (< 10,000' Above Field Elevation (AFE)) [28, 47]. The data to build the models are obtained from performance calculators developed by Boeing. Parametric equations are used to model the TSFC as a function of the ambient conditions, Mach number, altitude, and the net thrust. The model is referred to as the Senzig-Fleming-Iovinelli (SFI) model, after the names of the lead authors in [28, 47]. The SFI model is used in AEDT to model fuel burn in the terminal areas. Unlike the BADA model, the SFI model expands the set of inputs used to calculate the TSFC. Nonlinear equations are used to model the TSFC. However, like BADA, the SFI model needs thrust as an input to estimate the TSFC (and subsequently the fuel flow and fuel burn). The model also does not give uncertainty estimates for the fuel flow rate. Finally, the model is valid only for altitudes below 10,000' AFE.

## 2.4 Model refinements using operational data

Operational data on gate-to-gate fuel consumption, time spent on the ground and in the air, take-off weight, and origin-destination pairs have been used to refine aircraft performance models based on the BADA model as well as the ICAO databank. A 10% improvement in fuel burn estimates is achieved through the incorporation of such operational data [53]. Flight recorder data have been used to develop least squares regression models to estimate taxi fuel consumed as functions of taxi time, number of stops, and number of turns on the surface, as well as the taxi time and number of acceleration events during taxi [54]. Flight recorder data have also been used to enhance the System for assessing Aviation's Global Emissions (SAGE) [55], a precursor to AEDT [56]. In other work, multiple regression has been applied to FOQA data to identify anomalous outliers in fuel burn. Cruise phase fuel burn for each engine of the Boeing 757 is regressed on the calibrated speed, gross weight, and the engine RPM. The total fuel burn (from

all engines combined) is found to be higher than that given by the manufacturer's performance manual [57]. In summary, these studies illustrate the improvement in fuel burn estimation that could be achieved through the use of operational data.

## 2.5 Estimation of takeoff weight

The Takeoff Weight (TOW) of an aircraft is an essential parameter for modeling or estimating its trajectory and fuel consumption, as well as other aircraft performance characteristics, such as, its rate of climb/descent, range, endurance, ceiling, and takeoff distance [58, 33]. The models described in Section 2.3.2 further illustrate the need to know the TOW in order to accurately estimate fuel burn. However, TOW is not generally available outside the operating carrier, due to its dependence on proprietary information such as passenger load factors and operational strategies. The above facts motivate the development of models to estimate the TOW of a flight from accessible information.

Aircraft design studies have traditionally estimated the TOW by considering its components, namely, the payload weight, stage length fuel weight, operating empty weight, reserve fuel weight, and alternative fuel weight [59, 60, 10]. This approach is effective for studies in which the payload weight is an input. It can also be used to estimate the average TOW of an aircraft type over a set of operations for which the average passenger load factor is available [61]; for example, average passenger load factors for different origin-destination pairs are published in the United States by the Department of Transportation [62]. However, this method cannot be easily extended to estimate the TOW of a particular flight, as load factors of individual flights are not publicly known.

Prior studies have estimated the TOW for a particular flight using simulated or real aircraft trajectory information during the climb phase [63, 64, 65, 66]. They typically estimate an equivalent TOW such that the power in climb modeled using the equivalent TOW matches the energy rate observed on past trajectory points. The equivalent TOW is computed using either an adaptive mechanism, or least squares algorithms. Machine learning techniques have also been applied to radar data to estimate the TOW in order to predict the future aircraft trajectory

[67]. The methods proposed in these studies have been shown to be superior to the EUROCONTROL's Base of Aircraft Data (BADA) method for trajectory modeling [22]. However, due to the unavailability of ground truth data, the accuracies of these TOW estimates have not been evaluated. Instead, these models have been evaluated based on their trajectory prediction accuracy. Studies have also looked at developing non-iterative, closed-form, flight phase-based methods for TOW estimation by using aircraft performance data from BADA [33]. However, these TOW estimates are validated by comparison of the resultant range-payload diagrams with those supplied by the aircraft manufacturer, and not by direct comparison of the TOW estimates with ground-truth data. Models have also been developed which estimate the TOW based on the flight trip length. For the same aircraft type, longer distance flights are expected to have a higher TOW. One such model is the Aircraft Noise and Performance (ANP) model [34] used in AEDT. However, flight trip length is not a complete indicator of the TOW because of airline operations such as fuel tankering.

Recent work has used runway ADS-B data during takeoff to model the operational TOW, using analytical methods or methods based on least squares [68]. However, the resultant TOW estimates could not be validated due to the unavailability of ground truth data. Moreover, these studies assume no deration in the takeoff thrust, and a standard coefficient of friction for the ground roll, which result in approximate estimates of the operational TOW. Finally, phase-based models have been used for the Bayesian inference of TOW [69].

## 2.6 Discussion

A study of the prior literature illustrates the benefits of data-driven models and how they can help to overcome some of the limitations of simulation models, and reduce modeling assumptions. The literature survey also revealed insights into the desired characteristics of fuel burn models. In particular, statistical techniques help directly map input features (for example, the 4-D flight trajectory) to the desired outputs (for example, the aircraft fuel flow and fuel burn). These methods also enable quantification of uncertainty in the output arising from random disturbances and unmodeled features.

Most prior studies and models have used nonoperational data from ground tests, performance calculators, and flight manuals for model building. Results from such models may not be representative of real flight operations (as shown by the comparison between the ICAO Databank and FDR-derived estimates of fuel burn in Section 2.3.1). This observation motivates the use of operational flight data for model building. Prior studies also show that the BADA models, which are used in many trajectory modeling tools worldwide, do not adequately represent the factors which govern the fuel burn. The use of parametric (mostly linear) equations belies the complex nature of the mapping between trajectory and fuel burn. BADA also needs a thrust estimate to calculate the fuel burn. An improved modeling methodology should leverage a physical understanding of aircraft and engine behavior, in order to adequately account for different factors affecting fuel burn. It should also be able to map the trajectory to the fuel flow rate directly, bypassing the estimation of intermediate parameters (such as thrust). In other words, it should improve on existing architectures (for example, Figure 1-1), by making better use of operational data. The model architecture shown in Figure 1-2 illustrates such an alternative. We would also like models that reflect problem complexity, making nonparametric statistical techniques well-suited for this objective. We would also like a methodology that can model the fuel burn in all different phases of flight (that is, the surface, ascent, cruise, and descent).

Lastly, prior literature shows that an accurate estimate of TOW is necessary for accurately modeling aircraft performance (including its fuel burn). Therefore, the development of a validated methodology for estimation of takeoff weight along with its associated uncertainty is also an important research question.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Chapter 3

# Datasets

As mentioned in Chapter 1, this thesis develops a data-driven approach to model aircraft engine fuel burn. This chapter describes the different data sources and datasets used for model development.

## 3.1 Flight data recorder

Colloquially known as the 'black box', a Flight Data Recorder (FDR) is an instrument on board the aircraft which records the values of important parameters during flight. As a result, it is regarded as one of the most accurate sources of operational flight data. The FDR data in this thesis have been obtained from a major European carrier. The FDR dataset reports important aircraft and engine parameters such as trajectory, speeds, gross mass, fuel flow rate, pressure ratios, spool speeds, ambient pressure and temperature, aircraft attitudes, aircraft accelerations, positions of auxiliary devices, etc. as a function of time. The frequency of sample recording is higher in the rapidly changing phases of flight (such as takeoff, initial part of climb, latter part of descent, landing) and lower in those phases which undergo less rapid changes (such as cruise). There are two sets of FDR data in our possession, one set is from approximately 2006 (FDR-I) and the other set is from 2016 (FDR-II). The two FDR datasets have been procured at different times. It is important to note that being confidential in nature, the FDR data for any particular flight operation are not disseminated by the airline in general, thereby making it

impossible to obtain recorded fuel flow rate values for any particular flight. Therefore, there lies enormous promise in the ability to use the small subset of FDR data acquired in this study to develop fuel burn prediction models. The developed models are abstractions of real flight data and can be used to predict the fuel burn for a flight when its flight recorder data are unavailable. The FDR-I dataset has been used to develop models to predict the TOW and the fuel flow rates in the airborne phases of flight. The FDR-II dataset has been used to develop models to predict the fuel flow rates during taxi.

### 3.1.1 Aircraft types and engines

The FDR archives in possession contain data for 9 distinct aircraft/engine types. Only those flights are considered whose FDR records are uncorrupted, and whose trajectory can be clearly separated into distinct flight phases (like taxi, takeoff roll, ascent, cruise, descent, and touchdown) by employing simple trajectory-based criteria. The different flight phases and the criteria to identify them are elaborated in Section 3.1.2. The different aircraft and engine types included in the study are tabulated in Table 3.1 and Table 3.2. The approximate Maximum Takeoff Weight (MTOW), the approximate Operating Empty Weight (OEW), and the number of flights (# Flts.) are also shown. It should be noted that the MTOW and the OEW values in the tables are approximate, representative numbers for the aircraft types and have been obtained from the Jane's All the World's Aircraft database [70, 71, 72]. When not explicitly provided, the aircraft-engine match has been inferred using the JP Fleet database entries for the appropriate airline [73, 74].

Figure 3-1 shows histograms of the takeoff weights observed in the FDR-I dataset for different aircraft types, with the MTOW and OEW values overlaid. Figure 3-2 shows histograms of the total airborne fuel burn observed in the FDR-I dataset for different aircraft types.

### 3.1.2 Flight phase identification

Before analysis, the trajectory of each flight is split into the following different phases (Figure 3-3):

Table 3.1: FDR-I dataset: Aircraft types and engines.

| Aircraft Type | Engine Type | MTOW (kg) | OEW (kg) | # Flts. |
|---|---|---|---|---|
| A319-112 | 2×CFMI CFM56-5B6/2 or 2P | 64,000 | 40,160 | 130 |
| A320-214 | 2×CFMI CFM56-5B4/2 or P/2P | 73,500 | 42,100 | 169 |
| A321-111 | 2×CFMI CFM56-5B1/2 or 2P | 89,000 | 48,500 | 117 |
| A330-202 | 2×GE CF6-80E1A4 | 230,000 | 120,500 | 84 |
| A330-243 | 2×RR Trent 772B-60 | 230,000 | 120,600 | 100 |
| A340-541 | 4×RR Trent 553 | 372,000 | 170,900 | 52 |
| B767-300 | 2×GE CF6-80C2B7F | 156,490 | 86,955 | 91 |
| B777-300ER | 2×GE GE90-115B1 | 345,050 | 167,825 | 131 |

Table 3.2: FDR-II dataset: Aircraft types and engines.

| Aircraft Type | Engine Type | MTOW (kg) | OEW (kg) | # Flts. |
|---|---|---|---|---|
| A330-343 | 2×RR Trent 772B-60 | 230,000 | 124,600 | 182 |
| B777-300ER | 2×GE GE90-115BL | 345,050 | 167,825 | 127 |

1. Departure taxi (taxi-out)
2. Takeoff roll and wheels off
3. Ascent/Climb
4. Cruise
5. Descent
6. Touchdown
7. Arrival landing roll and taxi (taxi-in)

The methodology for identifying the different flight phases is as follows:

- The FDR records of aircraft trajectory parameters like latitude, longitude, pressure altitude, ground speed, and their derivatives with respect to time are used to identify the points of transition from one flight phase to the other.

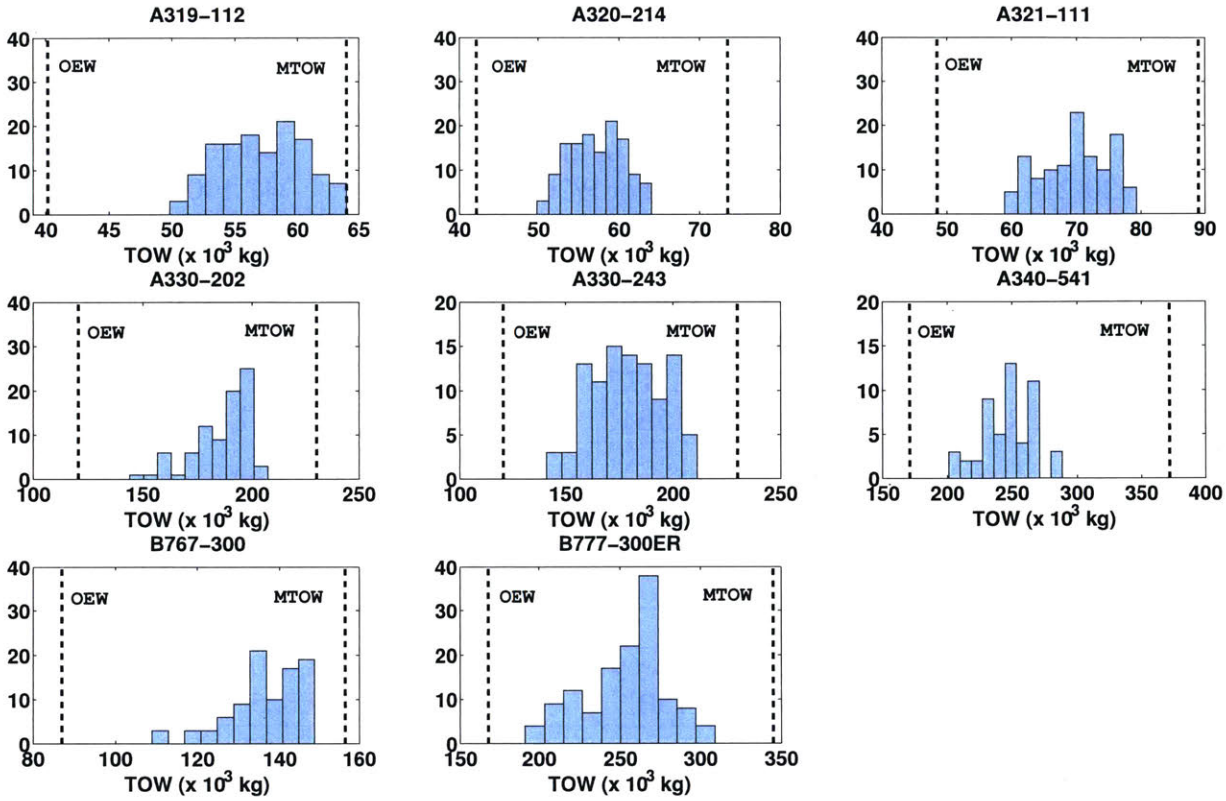



Figure 3-1: Histograms of the takeoff weights in FDR-I data with MTOW and OEW values overlaid.

- The phase from the first movement of the aircraft up to the start of the takeoff roll at the departure airport is the departure taxi (taxi-out) phase.
- The start of the takeoff roll is identified by a sudden increase in the aircraft acceleration post departure taxi. The aircraft then accelerates down the runway towards takeoff.
- The wheels-off condition (end of takeoff ground roll) is identified by the pressure altitude relative to the departure airport elevation becoming 0 and then subsequently increasing continuously and rapidly as the aircraft climbs.
- Climb follows, until the start of the cruise phase (top of climb).
- The start of cruise (top of climb) is identified by the beginning of leveling out of the altitude for an extended time period.

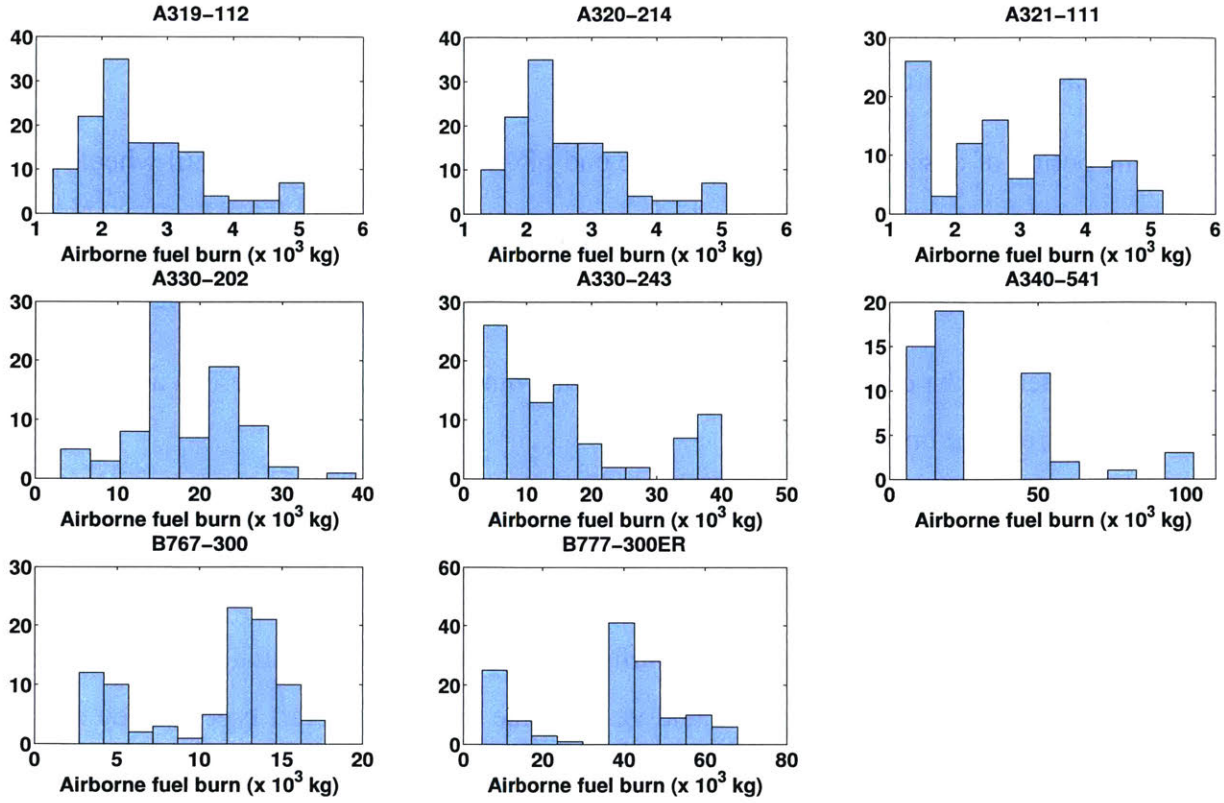



Figure 3-2: Histograms of the total airborne fuel burn in FDR-I data.

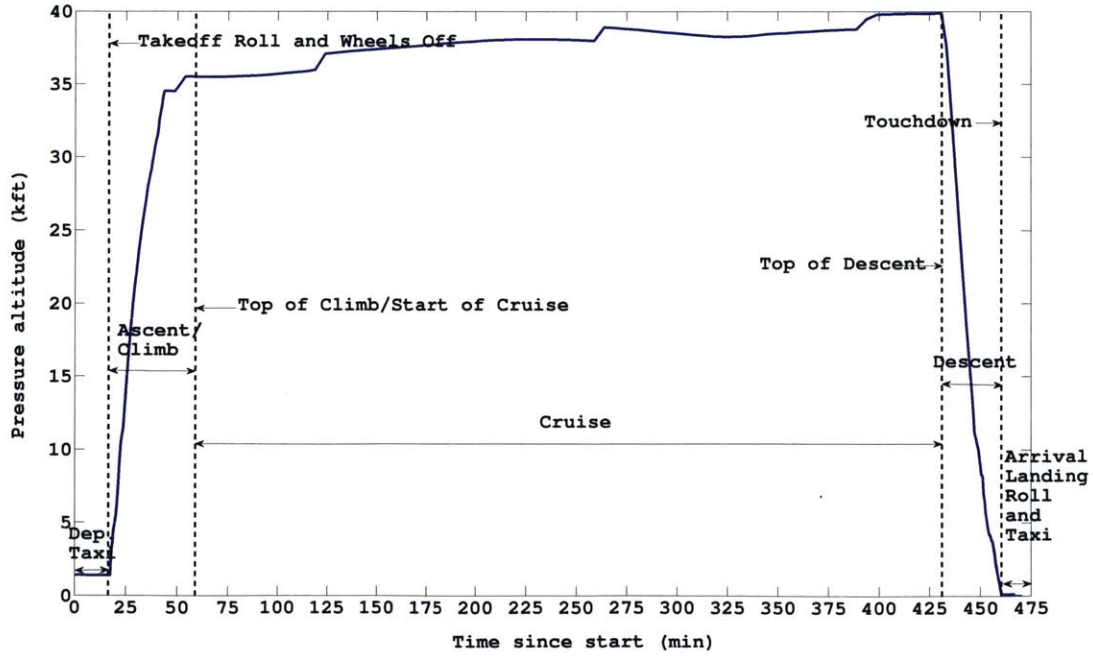



Figure 3-3: A typical profile of the pressure altitude versus time and the different flight phases (adapted from [49]).

- The end of cruise (top of descent) is identified by the beginning of a rapid and continuous decrease in aircraft altitude.
- The period of descent continues from the end of cruise (top of descent) until wheels-on.
- Touchdown (or wheels-on) is identified by the value of the pressure altitude relative to the arrival airport elevation becoming 0 (or below a certain threshold value).
- Subsequently, the entire phase until the aircraft comes to a stop at the arrival airport is the landing roll and taxi phase.
- It is important to note that all the above criteria make use of just the flight trajectory information as a function of time.
- The exact criteria for flight phase identification are adopted by manual inspection of the trends in the FDR trajectory data and vary for the different aircraft types. They can be easily tweaked, if required.

In addition to the main phases identified above, two sub-phases are further identified in this thesis: Climb out and approach. Climb out and approach are the parts of ascent and descent, respectively, occurring below 3,000 ft AFE. These phases are a part of the ICAO LTO cycle and considered separately as they are the airborne phases occurring the closest to the ground, thereby impacting air quality in the vicinity of the ground.

### 3.1.3 Airports

A total of 89 airports are found serving the different flights in the FDR archives. Figure 3-4 shows the geographical extent of the flight operations considered. The airport Above Mean Sea Level (AMSL) elevation varies from -11 ft to 5,558 ft.

## 3.2 Airport Surface Detection Equipment, Model X

The Airport Surface Detection Equipment, Model X (ASDE-X) is a surveillance system at major US airports which is used primarily for collision avoidance on the airport surface [75].


Figure 3-4: Airports in the FDR archives.

Using data from surface radars, terminal radars, Automatic Dependent Surveillance-Broadcast (ADS-B) sensors, and aircraft transponders, ASDE-X tracks all aircraft and ground vehicles on the airport surface and in the airspace within 8 km of the airport. In this study, ASDE-X data have been used to demonstrate the practical usability of the surface fuel burn and TOW prediction models when aircraft ground tracks are given. The ASDE-X archives in this study comprise records of 166 flights from 2016, which are inferred to correspond to the flights in the FDR-II dataset (based on a maximum likelihood match).

For a particular flight, its ASDE-X record contains track information such as latitude, longitude, speed, heading, and altitude, as functions of time. There is also information about the departure fix, aircraft identification, and aircraft type. Resolution of the trajectory variables in the ASDE-X dataset is as follows:

- Sampling rate: 1 Hz
- Latitude and longitude: $< 10^{-6}\,^{\circ}$
- Heading: $< 10^{-6}\,^{\circ}$
- Speed: 0.51 $\text{ms}^{-1}$ (1 knot)

A drawback of the ASDE-X archives is that the raw data are very noisy and have missing fields. These tracks, therefore, need to be smoothed before use, as described in Chapter 6.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Chapter 4

# Statistical Methods

This chapter gives a brief primer on the different statistical methods used for model development in this thesis. It gives only a high-level overview of the methods and readers are asked to refer to the cited texts for more details about the methods. The different methods described in this chapter are the Ordinary Least Squares (OLS) regression, methods based on regression trees (Classification and Regression Trees (CART) and Least Squares Boosting (LSB)), and Gaussian Process Regression (GPR).

## 4.1 Supervised learning

There are many real-life problems which involve the prediction of an output variable in response to an input variable. For example, one might be interested in predicting the weight of an individual (the output) given his/her height and age (the inputs). One might be interested in predicting if it will rain or not on a particular day (the output) given the temperature, pressure, and whether it rained or not the previous day (the inputs). Predicting the output requires the development of a mapping from the inputs to the outputs. One way to develop this mapping is by using data containing information about the input and the output variables of interest. For example, in the first problem, height, weight, age values for different people in a representative sample can be collected. In the second example, historical weather data can be collected. The values of the output and the input variables in the collected data can be used to develop a mapping between

the two and this mapping can then be used to predict the unknown output at a new input not seen before. This process of model creation using a subset of data containing values of both the output and the input variables is known as supervised learning [32] (as model creation is "supervised" by actual values of the output variable in the collected dataset).

Based on the nature of the output variable, the supervised learning problem can be of two different kinds. Developing a mapping to estimate the value of a continuous output (as in the first example) is known as a regression problem. Developing a mapping to predict which of a discrete set of classes the output falls in (as in the second example where there are just two classes - either it will rain or it will not rain) is known as a classification problem. In this thesis, the objective is to map the fuel flow rate (and the TOW) to aircraft trajectory variables. FDR data containing values of the fuel flow rates, the TOW, and the trajectory variables at every instant in flight are used for developing the mapping. Thus, fuel flow rate and TOW modeling as investigated in this thesis are supervised learning problems. More specifically, they are regression problems since fuel flow rate and TOW are continuous variables.

### 4.1.1 Regression

In regression, the output of the $i^{\text{th}}$ observation ($y_i$) is assumed to equal a function of the input features ($\mathbf{x_i}$) plus some noise ($\epsilon_i$),

$$y_i = f(\mathbf{x_i}) + \epsilon_i. \tag{4.1}$$

The goal is to estimate the underlying regression function $f(\mathbf{x})$ using a set of given inputs and their corresponding given outputs. Once the regression function is estimated, it can be used to estimate the unknown output corresponding to a new input ($\mathbf{x}^*$),

$$\widehat{y^*} = \hat{f}(\mathbf{x}^*). \tag{4.2}$$

The ultimate objective of any regression model is to provide accurate predictions of the output variable, given a new input vector. Prior to building any model using regression, the dataset is divided into three mutually exclusive parts:

- **Training dataset**: This dataset, containing the true values of the input and the output

variables, is used to train the regression model to estimate the regression function, thereby generating a mapping from the input to the output variables.

- **Validation dataset**: This dataset, separate from the training dataset, is used to evaluate a set of models developed in order to choose one model that gives the "best" predictive performance. The validation dataset is also known as the development dataset.
- **Test dataset**: This dataset, separate from both the training and the validation sets, is used to evaluate the predictive performance of the chosen "best" model.

Since the objective is to get accurate predictions on new inputs, it is important to evaluate a model for its generalized predictive performance on a dataset that has not been used for training. Evaluating a model on data that have been seen during training can lead to overly-optimistic estimates of the prediction accuracy.

## 4.2 Ordinary Least Squares regression

Ordinary Least Squares (OLS) is the most commonly used regression algorithm. It is a probabilistic, parametric regression method. The noise ($\epsilon_i$ in Equation 4.1) is assumed to follow a Gaussian distribution thereby lending an underlying probability distribution to the output. The regression function is assumed to be a linear combination of the appropriate input features [76]. The coefficients in this linear combination are the parameters, and are estimated via maximum-likelihood estimation.

$$y_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{k,i} + \epsilon_i, \text{ where } \epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_n^2) \tag{4.3}$$

Here, $\beta$s are the parameters to be estimated, $x_{k,i}$ is the $k^{\text{th}}$ component of the input feature vector at the $i^{\text{th}}$ observation ($\mathbf{x_i}$), $p$ is the number of input features, $\mathcal{N}(0, \sigma_n^2)$ is the Gaussian distribution with 0 mean and variance $\sigma_n^2$, and 'i.i.d.' stands for 'independent, identically distributed'.

As an example, let us consider an output/dependent/predicted variable $y$ to be modeled as a function of two input/independent/predictor variables $x_1$ and $x_2$. If linear basis functions are

adopted, then for each observation, Equation 4.3 becomes

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i. \tag{4.4}$$

The model input features are $x_1$ and $x_2$, and the output feature is $y$. If quadratic basis functions are adopted, then for each observation, Equation 4.3 becomes

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,i} + \beta_4 x_{1,i}^2 + \beta_5 x_{2,i}^2 + \epsilon_i. \tag{4.5}$$

The model input features are now $x_1$, $x_2$, $x_1 x_2$, $x_1^2$, and $x_2^2$, and the output feature is $y$. In OLS regression, this form of the input and output feature vectors needs to be assumed prior to regression and each input feature is associated with a regression parameter/coefficient which needs to be estimated.

Equation 4.3 can be written in matrix form as

$$\mathbf{y} = \mathbf{X_h}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{4.6}$$

Here, $\mathbf{y}$ is the $n \times 1$ vector of the output values corresponding to the $n$ observations in the training dataset, $\mathbf{X_h}$ is the $n \times (p+1)$ matrix of training input vectors with a leading column of ones added, $\boldsymbol{\beta}$ is the $(p+1) \times 1$ vector of the regression parameters to be estimated, and $\boldsymbol{\epsilon}$ is the $n \times 1$ vector of the noise values corresponding to the $n$ observations. $\mathbf{X_h}$ is given as

$$\mathbf{X_h} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{\mathrm{p},1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{\mathrm{p},2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,\mathrm{n}} & x_{2,\mathrm{n}} & \dots & x_{\mathrm{p,n}} \end{bmatrix}. \tag{4.7}$$

The least squares approach (which is equivalent to a maximum likelihood estimation for Gaussian noise) estimates the regression parameters as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X_h^T X_h})^{-1} \mathbf{X_h^T y}. \tag{4.8}$$

The mean prediction of the output at a new input (column) vector $\mathbf{x}^*$ is then given as

$$\widehat{y^*} = \mathbf{x_h^*}^T \hat{\boldsymbol{\beta}} \tag{4.9}$$

$$\text{where, } \mathbf{x_h^*} = \begin{bmatrix} 1 \\ \mathbf{x}^* \end{bmatrix}. \tag{4.10}$$

The predictive distribution for the predicted output in the OLS regression framework is a Student's t-distribution. The 95% prediction interval (0.95PI) for the output $y^*$ corresponding to the new input vector $\mathbf{x}^*$ is given by the 95% Highest Density Interval (HDI) [77] of this t-distribution[1]. The 95% prediction interval is an interval centered about the mean prediction ($\widehat{y^*}$) within which the true output falls with probability 0.95.

$$0.95\text{PI} = \{y^* | y^* \in [\widehat{y^*} - \mathcal{T}(0.975; n-p-1)s_{\text{pred}}, \tag{4.11}$$

$$\widehat{y^*} + \mathcal{T}(0.975; n-p-1)s_{\text{pred}}]\} \tag{4.12}$$

$$\text{where, } s^2_{\text{pred}} = \text{MSE}(1 + \mathbf{x_h^*}^T(\mathbf{X_h^T X_h})^{-1}\mathbf{x_h^*}), \text{ and} \tag{4.13}$$

$$\text{MSE} = \frac{(\mathbf{y} - \mathbf{X_h}\hat{\boldsymbol{\beta}})^\mathbf{T}(\mathbf{y} - \mathbf{X_h}\hat{\boldsymbol{\beta}})}{n-p-1} \tag{4.14}$$

$\mathcal{T}(0.975; n-p-1)$ refers to the $97.5^{\text{th}}$ percentile of the Student's t-distribution with $n-p-1$ degrees of freedom ($\zeta$), $s^2_{\text{pred}}$ is the estimated variance for the predicted output, and MSE is the regression mean squared error.

The residuals, given by $\mathbf{y} - \mathbf{X_h}\hat{\boldsymbol{\beta}}$, are the error estimates at the different observations in the training dataset. The underlying probabilistic assumptions on the noise terms require the residuals to follow certain properties. These properties, which can be used as diagnostic tools to check the quality of the OLS regression models, are as follows:

- The residuals, when plotted against the input features, should not show any specific trends. They should be located around zero and should be independent of the input feature values. A systematic variation of the residuals with the input features could indicate

[1] A 95% Highest Density Interval (HDI) is an interval in the domain of a probability distribution such that, (i) the probability mass within the interval is 0.95, and (ii) every point inside the interval has a probability density not less than every point outside it. The interval is unique and the shortest among all possible 95% confidence intervals for that distribution.

an improper choice of the basis functions used for regression.

- The residuals, when plotted against products of input features, should not show specific trends. Presence of trends could indicate that the interactions among the input features are not sufficiently captured in the model.

- The noise terms are assumed to have a constant variance ($\sigma_n^2$). Thus, the residuals should also have similar variance (homoskedasticity) at different levels of the input features or the fitted values. A non-constant variance of the residuals (heteroskedasticity) indicates model misspecification.

- The noise terms are assumed to be independent. Therefore, the residuals should be independent too. One way to check for residuals' independence is to plot the residuals against their lagged values. Systematic trends in this plot could indicate that the residuals are not independent of one another.

- The noise terms are assumed to be normally distributed. Therefore, the residuals should also be normally distributed. Normality of residuals can be checked using normal probability plots or through statistical testing. Deviation from normality indicates a major model misspecification as error normality is the basis for many model results (such as the distributions of the model coefficients, and model predictive distributions).

Remedial measures exist to correct for possible model misspecifications. One such remedial measure is feature transformation to change the basis functions (such as using quadratic instead of linear basis functions, taking logarithmic or square root transformations of the original features, etc.). More such remedial measures and details of the OLS regression can be found in [76]. In this thesis, inbuilt MATLAB®[78] toolboxes are used for building models using OLS regression.

The advantage of OLS regression is that it is simple to interpret, computationally inexpensive, gives a closed form solution for the parameters, and, being probabilistic, can generate the predictive distribution of the output at a new input. However, one needs to choose a form of the input feature vector ($\mathbf{x}$) (for example, polynomial, logarithmic, trigonometric, etc.) before

undertaking the parameter estimation. This poses a considerable challenge in problems without prior knowledge of the exact functional form of the feature vector (which may be nonlinear), and motivates the adoption of nonparametric methods which enable a data-driven tuning of the regression function without having to explicitly assume a form of the feature vector prior to regression.

## 4.3 Regression trees

As mentioned in the previous section, use of nonparametric methods frees the user from the choice of the form of the regression features prior to model building. One such nonparametric method is the use of regression trees.

Regression trees are decision trees used for the purpose of regression. Modeling using regression trees involves generating rules based on values of the input variables. Sequentially following these rules leads to the output prediction. The regression tree based methods, as used in this thesis, are nonprobabilistic (i.e., they do not assume an underlying probability distribution for the output). They are nonparametric as there are no specific parameters which need to be estimated in order to build the regression model. Two algorithms based on regression trees are investigated in this thesis: Classification and Regression Trees (CART) and Least Squares Boosting (LSB).

### 4.3.1 Classification and Regression Trees

Classification and Regression Trees (CART) is a popular algorithm developed by Leo Breiman et al. [79]. Starting with the root node containing all the data points, CART carries out recursive binary splitting of the data. The split criteria are always of the form $X < \chi$ where $X$ is a particular predictor variable and $\chi$ is the split point. Points in a node satisfying the split criterion go into the left child node and the others go into the right child node. This split at each node is a locally optimal split, chosen so as to maximally reduce the weighted sum of the mean squared errors of the resulting nodes. The sum is weighted by the fraction of the observations going in the left child and the right child nodes. The algorithm assigns the mean of

the values of the dependent variable in a node as the value of that node. The recursive splitting continues till the algorithm reaches a stopping criterion. The nodes which do not further split are known as leaf nodes. The algorithm stops splitting a node when either the improvement in mean squared error due to a further split drops below a threshold, or when a further split drops the number of observations in a node below a pre-defined value, or when splitting is no longer possible as the predictors have the same distribution for all the points in the node, or when only one observation remains in the node. In this thesis, a minimum number of 10 observations is required to be present in each node. In each leaf node, the algorithm assigns a constant value equal to the mean of the dependent variable for all the observations in that node. Each leaf node represents a partition of the input feature space which is assigned a constant value of the output variable.

The tree growing procedure normally produces a very deep tree which overfits the data. Hence, the tree is pruned as a means of regularization. Pruning re-combines leaves (to reduce the total number of leaves and hence, the tree complexity) in a manner which keeps the increase in training mean squared error to a minimum. This successive pruning produces a sequence of subtrees. The mean squared error given by each subtree on out-of-sample data is chosen as the metric of its generalized performance. In this thesis, 10 fold cross validation is used to calculate the out-of-sample mean squared error and its standard deviation. The most parsimonius subtree having its mean squared error within one standard deviation of the minimum error across all subtrees is chosen as the final tree to be used further for modeling and prediction of the fuel flow rate.

Figure 4-1 shows a typical regression tree built by CART and the partitions induced by it in the predictor space [80]. The tree is a toy model which regresses fuel flow rate per engine (in kg/s) on two variables: Altitude ($h$, in m) and ground speed ($V_{\mathrm{GS}}$, in m/s) for the A320-214 in ascent. For ease of representation and interpretability, only 1000 observations randomly chosen from the FDR dataset for the A320-214 are used for model building in Figure 4-1[2]. The tree clearly shows that prediction is based upon a set of simple decision rules which direct each

[2]It must be emphasized that the regression tree and the predictor space partitions in Figure 4-1 are solely for the purpose of understanding, via a toy problem, how a tree developed using CART works. The actual trees for fuel flow rate prediction using CART are built using a different set of features, and a larger number of observations (as explained in Chapter 5).

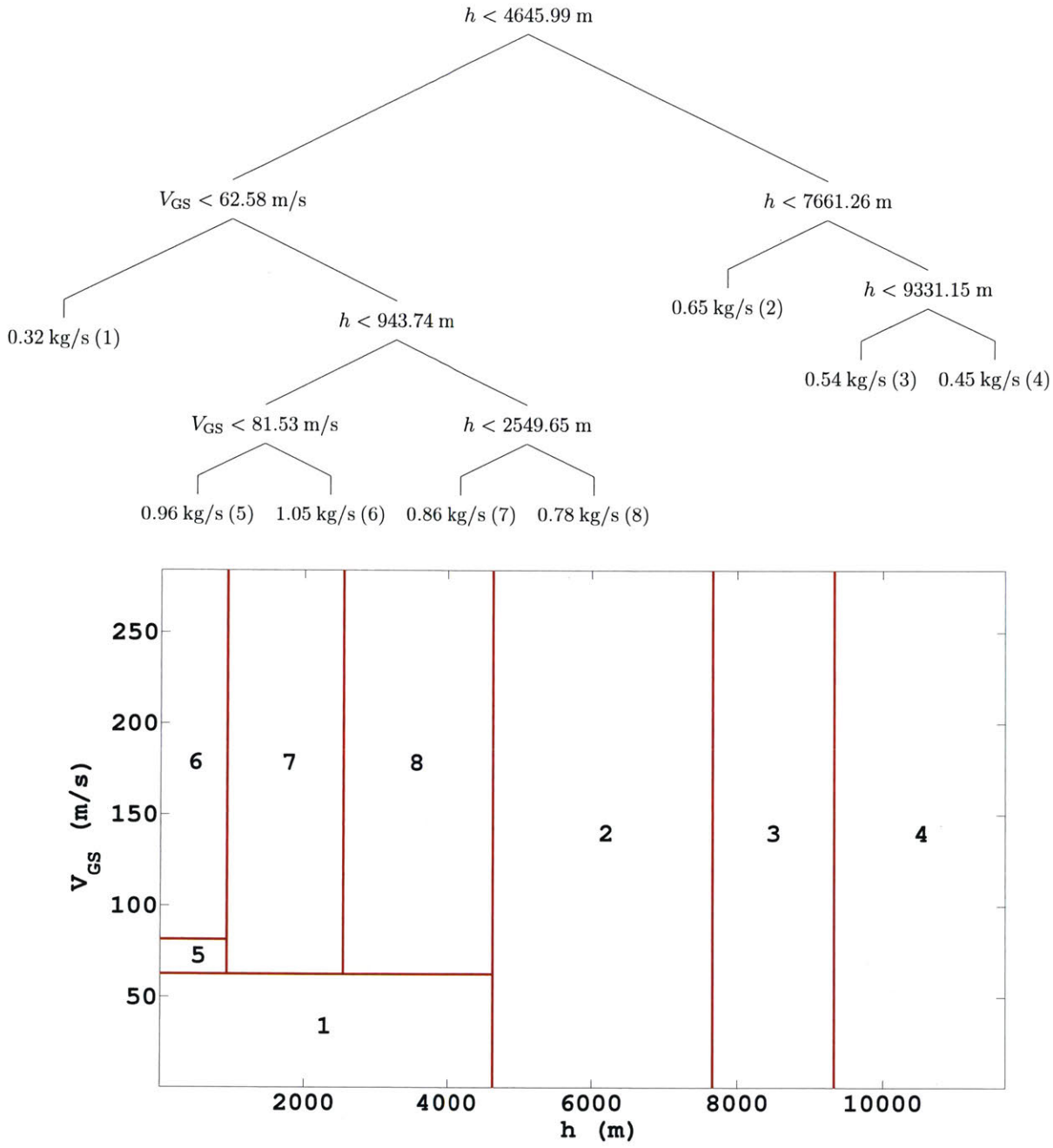



Figure 4-1: A320-214: CART visualization in ascent: (top) the tree, and (bottom) the partitions of the predictor space induced by the tree. The numbers inside the partitions correspond to the leaf node numbers in the tree (written in parentheses for each leaf node) (reproduced from [80]).

input vector to a leaf node and assign a node-specific value for the fuel flow rate. For example, one rule states that if the altitude is less than 4645.99 m and the ground speed is less than 62.58 m/s, then the fuel flow rate per engine is 0.32 kg/s (partition 1 in Figure 4-1). This simplicity

enables fast prediction. It is also easy to see how the tree seamlessly captures interactions among predictors. This is evident from the fact that splitting variables seen in the left subtree of a node could be different from those seen in the right subtree. The tree breaks up the predictor space into different partitions, each corresponding to one leaf node. Each partition represents that region of the predictor space where all points show homogeneity in fuel flow rate values. Thus, instead of fitting a global model to the entire dataset, CART identifies homogeneous regions of data and fits a separate model in each region. This has the capability of improving model accuracy as well as capturing the variability in data properly.

There are several advantages of using CART [32]. CART is a fast, computationally inexpensive, and a relatively automatic algorithm, not requiring much intervention from the user. It can be easily scaled to large datasets. The trees generated are easily interpretable. CART is a nonparametric algorithm (as there are no regression parameters present, unlike, for example, the regression coefficients in ordinary least squares regression). Being a nonparametric method, the problem of having to choose the form of the input features prior to model building does not arise. The trees also capture interaction among different predictors (up to the depth of the tree). Regression trees developed by CART are also robust to outliers in the input data and can deal with irrelevant data. The CART algorithm can easily handle missing values in the data. Lastly, the algorithm is invariant to monotonic transformations of the inputs.

### 4.3.2 Least Squares Boosting

The CART algorithm generates only one tree, which may be unstable and can have low predictive power. A combination of several such 'weak' regression trees can be expected to yield models with better prediction capability. Boosting [32] is an ensemble method which combines several 'weak' learners to yield a 'stronger' model. In this thesis, each weak learner is a regression tree generated by CART (as explained in Section 4.3.1). The boosting algorithm is run for many ensemble cycles. Based on the minimization of the cross validated mean squared error over the training dataset, the number of ensemble cycles in this thesis is 100. The algorithm starts by building a CART-based weak learner to the entire dataset. In each successive cycle, the boosting algorithm fits a weak tree learner to the residuals from the previous cycle. The

objective is to minimize the squared error loss and hence, the name Least Squares Boosting (LSB). The output predictions at a particular input from all the cycles are linearly combined to give the model prediction at that input. The contribution from each cycle is controlled by a manually set learning rate which, in this thesis, is set to a value of 0.1. The learning rate thus, shrinks the contribution from each tree.

The advantage of the LSB algorithm lies in its superior predictive ability compared to a single CART model. However, by combining many single CART models, the algorithm loses the interpretability seen in a single tree. The LSB algorithm is also slower than the CART algorithm. Both the CART and the LSB algorithms have seen widespread application in diverse areas like solar radiation modeling [81], strength modeling of composites [82], modeling of signalling traffic in mobile networks [83], highway safety studies [84], health monitoring of engineering components [85], pollution studies [86], and ecological studies [87].

### 4.3.3 Bootstrapping

For nonprobabilistic methods (such as CART and LSB) which do not assume an underlying probability distribution on the output, the 95% prediction intervals (used to quantify uncertainty in the output prediction) can be developed using the method of bootstrapping. Bootstrapping starts with re-sampling the training dataset with replacement. Each such sample set contains the same number of observations as in the original training dataset. Many such samples sets are created (1000 in this thesis) and each set is used for training a model. This model gives one value of the output prediction at a new input vector. Different sample sets lead to different models trained which, in turn, lead to different predictions at the same input. Thus, at every new input, a sample of output predictions is obtained. This sample can be used to develop 95% prediction intervals for the predicted output at the new input (for example, by taking the inter-percentile range between the $2.5^{\text{th}}$ and the $97.5^{\text{th}}$ percentiles of the prediction samples). More details about the bootstrap method can be found in [88]. Bootstrapping can be computationally expensive and thus, computing prediction intervals for nonprobabilistic methods can be challenging. In this thesis, MATLAB® toolboxes are used for running the CART, LSB, and the bootstrap algorithms [78].

## 4.4 Gaussian Process Regression

From the discussion in the previous sections, it is apparent that a regression method is desired that is both nonparametric (i.e., does not assume a form of the regression function prior to model building) and probabilistic (i.e., gives a predictive distribution of the output variable which can yield prediction intervals easily without having to resort to computationally challenging algorithms such as the booststrap method). One such method is Gaussian Process Regression (GPR), which has found application in diverse areas, including biomedical applications and health care [89, 90, 91], remote sensing [92, 93], music [94], robotics [95], cellular communications [96], and material microstructure analysis [97].

Gaussian Process Regression (GPR) uses Gaussian Processes for regression. A function $f(\mathbf{x})$ is said to follow a Gaussian Process (GP) if the function values at any finite set of inputs $\mathbf{x}$ follow a joint Gaussian distribution [98].

$$f(\mathbf{x}) \sim \mathcal{GP}(m_e(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{4.15}$$

where, $m_e(\mathbf{x})$ is known as the mean function and $k(\mathbf{x}, \mathbf{x}')$ is the kernel function over two inputs $\mathbf{x}$ and $\mathbf{x}'$, which governs the covariance among function values.

$$\mathrm{cov}(f(\mathbf{x}), f(\mathbf{x}')) = k(\mathbf{x}, \mathbf{x}') \tag{4.16}$$

GPR is a Bayesian statistical technique [77]. In a Bayesian setting, the parameter to be estimated is first given a 'prior' distribution which represents beliefs about the parameter prior to making any observations. Each observation made contributes to updating our belief about the parameter. The updated belief after the observation is made is represented by the 'posterior' distribution of the parameter. Mathematically, the Bayesian setting can be represented using the Bayes theorem,

$$p(\psi|\mathbf{y}) \propto p(\psi)p(\mathbf{y}|\psi) \tag{4.17}$$

where, $\psi$ is the parameter to be estimated, $\mathbf{y}$ is the vector of observations, $p(\psi)$ is the prior probability distribution of the parameter before making any observations, $p(\mathbf{y}|\psi)$ is the likelihood

of the observations given the parameter, and $p(\psi|\mathbf{y})$ is the posterior probability distribution of the parameter after making the observations.

In GPR, the regression function $\mathbf{f}(\mathbf{x})$ in Equation 4.1 is given a GP prior with a zero mean function. The noise parameter in Equation 4.1 is assumed to follow a Gaussian distribution $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_n^2)$. Under the assumption of a GP prior over the regression function and Gaussian noise, the output variable $y$ also follows a GP prior with a zero mean function and a noisy kernel function $k_{\text{noisy}}(\mathbf{x_p}, \mathbf{x_q})$ over $d$-dimensional column input vectors $\mathbf{x_p}$ and $\mathbf{x_q}$.

$$y \sim \mathcal{GP}(0, k_{\text{noisy}}(\mathbf{x_p}, \mathbf{x_q})) \tag{4.18}$$

The noisy kernel function over two inputs ($k_{\text{noisy}}(\mathbf{x_p}, \mathbf{x_q})$) relates to the kernel function for the GP prior over the regression function ($k(\mathbf{x_p}, \mathbf{x_q})$) (Equations 4.15–4.16) as

$$k_{\text{noisy}}(\mathbf{x_p}, \mathbf{x_q}) = k(\mathbf{x_p}, \mathbf{x_q}) + \sigma_n^2 \delta_{\text{pq}} \tag{4.19}$$

where, $\delta_{\text{pq}}$ is the Kronecker delta between the $p^{\text{th}}$ and the $q^{\text{th}}$ input vectors.

### 4.4.1 Kernel functions

A large amount of the flexibility in GPR results from the rich suite of kernel functions available to model the covariance among the function values. Any function which can give a valid (positive semi-definite and symmetric) covariance matrix over function values is a viable kernel function. By controlling the covariance between function values, these kernel functions control the smoothness of the functions. Some commonly used kernels are as follows [99, 98]:

- Squared Exponential kernel: This kernel function is a stationary kernel function, which depends on the inputs only via the Euclidean distance between them. It is used to model very smooth functions. It is given by

$$k_{\text{SE}}(\mathbf{x_p}, \mathbf{x_q}) = \sigma_f^2 \exp\left( -\frac{1}{2} \sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{{\ell_i}^2} \right). \tag{4.20}$$

- Exponential kernel: This stationary kernel function is used to model rough functions. It is given by

$$k_{\mathrm{E}}(\mathbf{x_p}, \mathbf{x_q}) = \sigma_f^2 \exp\left( -\sqrt{\sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{{\ell_i}^2}} \right). \tag{4.21}$$

- Matérn class of kernels: This class of stationary kernels models smoothness levels between those of the exponential and the squared exponential kernels. Two commonly used kernel functions from the Matérn class of kernels are as follows:

$$k_{\mathrm{Mat.},3/2}(\mathbf{x_p}, \mathbf{x_q}) = \sigma_f^2 \left(1 + \sqrt{3}r\right) \exp(-\sqrt{3}r) \tag{4.22}$$

$$k_{\mathrm{Mat.},5/2}(\mathbf{x_p}, \mathbf{x_q}) = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{5r^2}{3}\right) \exp(-\sqrt{5}r) \tag{4.23}$$

$$\text{where, } r = \sqrt{\sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{{\ell_i}^2}} \tag{4.24}$$

In Equations 4.22–4.23, the '3/2' and '5/2' refer to values of a parameter in the Matérn class of kernels which controls the level of smoothness. 3/2 and 5/2 are two values commonly used. In Equations 4.20–4.24, $\sigma_f^2$ is a variance parameter governing the magnitude of the kernel, $\boldsymbol{\ell}$ is the $d$-dimensional vector of length scales (one for each input dimension), and the subscript $i$ refers to the $i^{\text{th}}$ component of the vector. These kernel parameters are referred to as *hyperparameters* in GPR. Thus, the hyperparameter vector for the squared exponential, exponential, and Matérn kernels is $[\sigma_f^2\ \boldsymbol{\ell}]^T$.

- Dot product kernel: This is a non-stationary kernel, whose value depends on the actual location of the inputs (instead of the distance between them). It is given by

$$k_{\mathrm{DP}}(\mathbf{x_p}, \mathbf{x_q}) = \sigma_0^2 + {\mathbf{x_p}}^T \boldsymbol{\Sigma} \mathbf{x_q}, \tag{4.25}$$

$$\text{where, } \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2). \tag{4.26}$$

Here, $\sigma_0^2$ is the constant variance parameter, $\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2$ are the variance parameters

for each of the $d$ input dimensions, and 'diag()' is the diagonal matrix with its arguments constituting the diagonal entries. The hyperparameter vector for the dot product kernel is $[\sigma_0^2\ \sigma_1^2\ \sigma_2^2 \ldots\ \sigma_d^2]^T$.

Numerous other kernel functions exist, details of which can be found in [99]. The sum of two kernel functions is also a valid kernel function. In this thesis, additive kernels made up by summing a dot product kernel and different stationary kernels are used for GPR. Such kernels depend both on the location of the input vectors and the distance between them (so that the function values at two input vectors which are close together are correlated). The following kernels are used in this thesis for GPR:

- Dot Product Squared Exponential (DPSE) kernel: This kernel, formed by adding the dot product (Equation 4.25) and the squared exponential (Equation 4.20) kernels, is given by

$$k_{\mathrm{DPSE}}(\mathbf{x_p}, \mathbf{x_q}) = k_{\mathrm{DP}}(\mathbf{x_p}, \mathbf{x_q}) + k_{\mathrm{SE}}(\mathbf{x_p}, \mathbf{x_q}). \tag{4.27}$$

- Dot Product Exponential (DPE) kernel: This kernel, formed by adding the dot product (Equation 4.25) and the exponential (Equation 4.21) kernels, is given by

$$k_{\mathrm{DPE}}(\mathbf{x_p}, \mathbf{x_q}) = k_{\mathrm{DP}}(\mathbf{x_p}, \mathbf{x_q}) + k_{\mathrm{E}}(\mathbf{x_p}, \mathbf{x_q}). \tag{4.28}$$

- Dot Product Matérn (DPM) kernels: These kernels, formed by adding the dot product (Equation 4.25) and the Matérn class (Equations 4.22–4.23) of kernels, are given by

$$k_{\mathrm{DPM},3/2}(\mathbf{x_p}, \mathbf{x_q}) = k_{\mathrm{DP}}(\mathbf{x_p}, \mathbf{x_q}) + k_{\mathrm{Mat.},3/2}(\mathbf{x_p}, \mathbf{x_q}). \tag{4.29}$$

$$k_{\mathrm{DPM},5/2}(\mathbf{x_p}, \mathbf{x_q}) = k_{\mathrm{DP}}(\mathbf{x_p}, \mathbf{x_q}) + k_{\mathrm{Mat.},5/2}(\mathbf{x_p}, \mathbf{x_q}). \tag{4.30}$$

### 4.4.2 Hyperparameter inference

The next step after choosing the kernel function involves inference of the hyperparameters. Under GPR, the log posterior probability of the hyperparameter vector $\boldsymbol{\theta_h}$, given the matrix of

input vectors $\mathbf{X}$ and the vector of dependent/output variable values $\mathbf{y}$, is written as

$$\log p(\boldsymbol{\theta_h}|\mathbf{y}, \mathbf{X}) = \log p(\boldsymbol{\theta_h}) - \frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y} - \frac{1}{2}\log|\mathbf{K}_{\mathbf{y}}| - \frac{n}{2}\log(2\pi). \tag{4.31}$$

Here, $p(\boldsymbol{\theta_h})$ is the prior on the hyperparameter vector. The remaining terms in the right hand side of the above equation form the type-II log likelihood. $n$ is the number of observations in the training dataset used for model building, $\mathbf{X}$ is the $n \times d$ matrix of $d$-dimensional training inputs, $\mathbf{y}$ is the $n \times 1$ vector of the training output values, $\boldsymbol{\theta_h}$ is the vector of hyperparameters and $\mathbf{K_y}$ is the $n \times n$ covariance matrix derived from the noisy kernel function (Equation 4.19) over pairs of input vectors in the training set. $|\cdot|$ refers to the determinant of the matrix. The Maximum A Posteriori (MAP) estimate ($\widehat{\boldsymbol{\theta}_{\mathbf{h}}}$) found by maximizing the log posterior probability in Equation 4.31 is taken to be the estimate of the hyperparameter vector.

### 4.4.3 Prediction using Gaussian Process Regression

Once the hyperparameters are inferred, we wish to determine the predictive distribution of the output values $\mathbf{y}^*$ at a set of new inputs $\mathbf{X}^*$. This predictive distribution, which is also a Gaussian distribution in the GPR framework, is given as follows:

$$\mathbf{y}^*|\mathbf{X}^*, \mathcal{D} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{C}) \tag{4.32}$$

$$\boldsymbol{\mu} = \mathbf{K}(\mathbf{X}^*, \mathbf{X})\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y} \tag{4.33}$$

$$\boldsymbol{C} = \mathbf{K}(\mathbf{X}^*, \mathbf{X}^*) - \mathbf{K}(\mathbf{X}^*, \mathbf{X})\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{K}(\mathbf{X}^*, \mathbf{X})^{\mathbf{T}} + {\sigma_n}^2\mathbf{I}_{\mathbf{n}^*} \tag{4.34}$$

Here, $n^*$ is the number of new inputs at which predictions are desired, $\mathbf{X}^*$ is the $n^* \times d$ matrix of the set of new inputs, $\mathcal{D}$ is the set of training inputs and output values, $\boldsymbol{\mu}$ is the mean and $\boldsymbol{C}$ is the covariance matrix of the multivariate Gaussian predictive distribution, $\mathbf{K}(\mathbf{X}^*, \mathbf{X})$ is the $n^* \times n$ covariance matrix derived from the noisy kernel function (Equation 4.19) over pairs of new and training input variables, $\mathbf{K}(\mathbf{X}^*, \mathbf{X}^*)$ is the $n^* \times n^*$ covariance matrix derived from the noisy kernel function over pairs of the new input variables, and $\mathbf{I}_{\mathbf{n}^*}$ is the $n^* \times n^*$ identity matrix. The form of the predictive distribution highlights the advantage of using GPs - the

distribution is normal. The mean prediction is the mean of the Gaussian predictive distribution ($\boldsymbol{\mu}$) given by the GPR model. Along with mean predictions, we also get uncertainty estimates, in a mathematically tractable fashion. For a Gaussian predictive distribution over an output $y^*$ with mean prediction $\hat{y^*}$ and standard deviation of the predictive distribution $s_{\text{pred}}$, the 95% prediction interval is given by

$$0.95\text{PI} = \{y^* | y^* \in [\hat{y^*} - 1.96 s_{\text{pred}}, \hat{y^*} + 1.96 s_{\text{pred}}]\}. \tag{4.35}$$

Figure 4-2 shows a pictorial way to understand the essence behind GPR. Figure 4-2(a) shows the output function values drawn from the GP prior before making any observations. Based on the GP prior, several different functions are deemed possible. Figure 4-2(b) shows that as observations used for model building are taken into account, the space of the possible output functions restricts to include only those functions which can fit the observations (with the noise properly accounted for), and one is left with a smaller set of acceptable functions. The gray band in the figure shows the confidence intervals for the functions. The band is the narrowest around the observations (as the output values at such points are known with the greatest certainty). Thus, GPR 'tunes' the space of acceptable functions with the help of the observations.

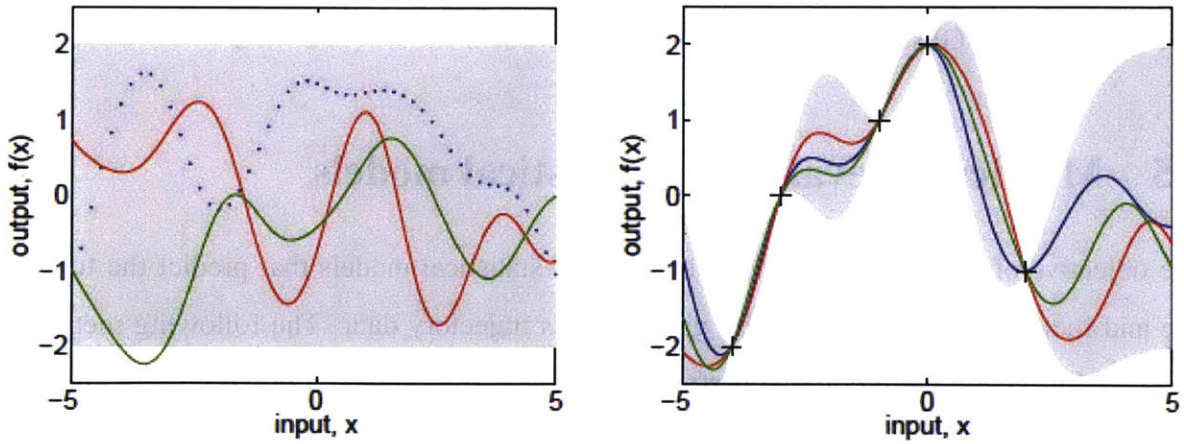



Figure 4-2: Visualization of Gaussian Process Regression, (a) the prior function space, and (b) the posterior function space (reproduced from [98]).

### 4.4.4 Approximate Gaussian Process inference and prediction

The exact GP inference and prediction explained in Sections 4.4.2–4.4.3 involves inverting an $n \times n$ matrix ($\mathbf{K_y}$), which requires $\mathcal{O}(n^3)$ operations. Hence, exact GP can be computationally very expensive for large datasets. In order to perform inference and prediction on large datasets, many approximations have been developed. One such approximation, used in this thesis, is the Fully Independent Conditional (FIC) approximation. In this approximation, $m_s$ artificial inputs, called the inducing inputs are introduced. The GPR function values at these inducing inputs are called inducing variables. Addition of these inducing variables enables us to make the first approximation that the vector of function values at the training inputs and the vector of function values at the new inputs where predictions are desired, are conditionally independent given the inducing variables. Additionally, in the FIC approximation, the training function values are assumed fully independent among themselves, given the inducing variables. The function values at the new prediction inputs are also assumed fully independent among themselves, given the inducing variables. Thus, all the function values at the training and the new inputs are considered fully independent, given the inducing variables. These approximations lead to massive computational savings if $m_s$ is much smaller than $n$. More details about the FIC (and other approximations) can be found in [100]. All GPR analysis in this thesis is conducted using the MATLAB®-based GPstuff toolbox [78, 101].

## 4.5 Metrics for evaluating statistical models

The objective of this thesis is the development of statistical models that predict the fuel flow rate and fuel burn profile for a flight, given just its trajectory data. The following metrics are developed to assess the predictive performance of a statistical model on an unseen prediction set (not used for model training), in terms of the accuracy of both the point and the interval estimates of the predicted output variable:

- **Mean Error (ME)**: The Mean Error (ME) is the mean of the values of the relative pre-

diction errors on the unseen prediction dataset.

$$\text{ME} = \frac{1}{n^*} \sum_{i=1}^{n^*} \left( \frac{\hat{y}_i - y_{i,\text{true}}}{y_{i,\text{true}}} \right) \tag{4.36}$$

Here, $n^*$ is the number of observations in the unseen prediction set, $y_{i,\text{true}}$ is the actual value of the output variable (ground truth) and $\hat{y}_i$ is the model point (mean) prediction of the output variable for the $i^{\text{th}}$ observation in the prediction set. The ME reflects whether the modeled point estimate over-predicts or under-predicts the ground truth. A negative value of ME indicates that the model point predictions, on the average, under-predict the ground truth. A positive value of ME indicates that the model point predictions, on the average, over-predict the ground truth. Thus, the ME is an indicator of the bias in the model predictions. It does not reflect the accuracy of the model point predictions.

- **Mean Absolute Error (MAE)**: The Mean Absolute Error (MAE) is the mean of the absolute values of the relative prediction errors on the unseen prediction dataset.

$$\text{MAE} = \frac{1}{n^*} \sum_{i=1}^{n^*} \left| \frac{\hat{y}_i - y_{i,\text{true}}}{y_{i,\text{true}}} \right| \tag{4.37}$$

The MAE reflects the accuracy of the model point prediction. A model with a low MAE is desired (with zero being the lowest possible value).

- **Prediction Coverage (PC)**: The Prediction Coverage (PC) is the fraction of observations in the unseen prediction set for which the ground truth values of the output variable fall within the 95% prediction intervals of the predicted output. The PC across all the observations in the prediction dataset is given by

$$\text{PC} = \frac{\sum_{i=1}^{n^*} \mathbb{1}_{0.95\text{PI}_i}(y_{i,\text{true}})}{n^*} \tag{4.38}$$

where, $\mathbb{1}$ is the indicator function, and $0.95\text{PI}_i$ is the 95% prediction interval for the $i^{\text{th}}$ observation in the prediction set. The PC indicates how well the prediction intervals

capture the variability of the output variable. A PC value close to 95% indicates that the model has been properly specified and formulated.

- **Predictive Log Likelihood (PLL)**: For a probabilistic model (such as GPR), PLL calculates the log of the likelihood of the actual output value (ground truth) in the unseen prediction set under the predictive distribution given by the model. The higher the PLL is, the better the actual output value can be explained by the model predictive distribution. For a Gaussian predictive distribution (with mean $\hat{y}_i$ and standard deviation $s_{\text{pred}_i}$ for the $i^{\text{th}}$ observation in the prediction set), the PLL across all the observations in the prediction set is given by

$$\text{PLL} = \sum_{i=1}^{n^*} \left[ -\frac{(y_{i,\text{true}} - \hat{y}_i)^2}{2s_{\text{pred}_i}^2} - \frac{1}{2}\log s_{\text{pred}_i}^2 - \frac{1}{2}\log(2\pi) \right]. \tag{4.39}$$

For a Student's t-distribution as the predictive distribution (as seen for OLS regression in Equations 4.11–4.14), the PLL across all the observations in the prediction set is given by

$$\text{PLL} = \sum_{i=1}^{n^*} \left[ \log\left\{\Gamma\left(\frac{\zeta+1}{2}\right)\right\} - \log\left\{\sqrt{\zeta\pi}\,\Gamma\left(\frac{\zeta}{2}\right)\right\} - \left(\frac{\zeta+1}{2}\right)\log\left(1+\frac{(y_{i,\text{true}} - \hat{y}_i)^2}{\zeta s_{\text{pred}_i}^2}\right)\right]. \tag{4.40}$$

Here, $\Gamma()$ refers to the Gamma function, $\zeta$ is the number of degrees of freedom for the t-distribution, $\hat{y}_i$ is the mean prediction and $s_{\text{pred}_i}$ is the estimated standard deviation for the $i^{\text{th}}$ observation in the prediction set.

- **Normalized Length of Prediction Interval (NLPI)**: This is the mean of the length of the 95% prediction intervals expressed as a fraction of the point estimate. For Gaussian predictive distributions, it is given by

$$\text{NLPI} = \frac{1}{n^*}\sum_{i=1}^{n^*} \frac{3.92 s_{\text{pred}_i}}{\hat{y}_i}. \tag{4.41}$$

For a Student's t-distribution as the predictive distribution, the NLPI is given by

$$\text{NLPI} = \frac{1}{n^*} \sum_{i=1}^{n^*} \frac{2\mathcal{T}(0.975; \zeta) s_{\text{pred}_i}}{\hat{y}_i}. \tag{4.42}$$

The NLPI indicates the extent of the relative uncertainty present in the estimated output.

## 4.6 Comparison of statistical models

In this thesis, the metrics developed in Section 4.5 are used to evaluate model predictive performance for the desired output variable on each flight in the unseen prediction set (not used for model training). It is often desired to compare these metrics among different models to determine which model has the best predictive performance. One way of such a model comparison is to compare the mean (or median) of the evaluation metric for the different models across all the flights. However, such a comparison does not account for the standard deviation seen in the metric across different flights. Thus, in this thesis, the metrics for different models are compared using statistical testing. This section gives a very brief overview of the different statistical tests used in this thesis. Readers are encouraged to refer to the in-text references cited for a more detailed exposition about these tests.

Let $\upsilon$ be the metric which needs to be compared for two different models $M_i$ and $M_j$. The null hypothesis ($\mathcal{H}_0$) and the alternate hypothesis ($\mathcal{H}_A$) for a two-sided statistical comparison are

$$\begin{aligned} \mathcal{H}_0 &: \upsilon_{M_i} - \upsilon_{M_j} = 0, \\ \mathcal{H}_A &: \upsilon_{M_i} - \upsilon_{M_j} \neq 0. \end{aligned} \tag{4.43}$$

Since two different models are being tested on the same set of flights, the statistical testing is a case of testing using repeated measures (or "paired" samples). A Wilcoxon signed-rank test [102] is used as a nonparametric statistical test in this thesis for pairwise model comparison. The Wilcoxon test takes as input the vector of $\upsilon_{M_i} - \upsilon_{M_j}$ for corresponding flights in the sample. Being nonparametric, the test does not require any assumption on the distribution from which

$v_{M_i} - v_{M_j}$ are drawn. The test output is a test statistic which, for a set of more than 10 flights, approximately follows a standard normal distribution under the null hypothesis. The p-value for the resulting test statistic is then determined to be the probability of observing a value as, or more, extreme than the calculated test statistic under the standard normal distribution (in both the left and the right tails of the distribution). For a level of significance $\alpha_s$, the null hypothesis is rejected if the p-value is less than $\frac{\alpha_s}{2}$. The null hypothesis cannot be rejected with the given data if the p-value is greater than or equal to $\frac{\alpha_s}{2}$. In this thesis, the value of $\alpha_s$ is taken to be 0.05.

In some cases, a one-sided statistical comparison may be desired. For a left-tailed comparison, the null and alternate hypotheses are

$$\begin{aligned} \mathcal{H}_0 &: v_{M_i} - v_{M_j} \geq 0, \\ \mathcal{H}_A &: v_{M_i} - v_{M_j} < 0. \end{aligned} \tag{4.44}$$

For a right-tailed comparison, the null and alternate hypotheses are

$$\begin{aligned} \mathcal{H}_0 &: v_{M_i} - v_{M_j} \leq 0, \\ \mathcal{H}_A &: v_{M_i} - v_{M_j} > 0. \end{aligned} \tag{4.45}$$

A similar Wilcoxon signed-rank test can be conducted for one-sided statistical comparison. For a left (right)-tailed comparison, the p-value for the resulting test statistic is determined to be the probability of observing a value as, or more, extreme than the calculated test statistic under the standard normal distribution (in the left (respectively, right) tail of the distribution). For a level of significance $\alpha_s$, the null hypothesis is rejected if the p-value is less than $\alpha_s$.

In some cases, more than one statistical comparisons may be simultaneously conducted. For example, the evaluation metric for model $M_i$ may be compared with that from both model $M_j$ and model $M_l$ simultaneously. In such cases of multiple pairwise comparisons, the p-value for each test should be compared with the significance level corrected with a Bonferroni correction factor [76]. This correction ensures that the family significance level for all the tests together is at most $\alpha_s$. For $k$ number of simultaneous tests, the effective level of significance for each

test in the family of tests under the Bonferroni correction is $\frac{\alpha_s}{2k}$ for two-sided testing and $\frac{\alpha_s}{k}$ for one-sided testing.

Finally, in some cases, it may be desirable to compare $k$ models for all pairwise combinations. Thus, there are $\frac{k(k-1)}{2}$ statistical comparisons to be simultaneously performed. The Wilcoxon signed-rank test can be performed and the p-value for each test can be compared to the significance level modified by the Bonferroni correction, as explained above, in the absence of the assumption of any underlying parametric distribution for the differences in the evaluation metrics. However, if a Gaussian parametric distribution is found to hold for the differences in the evaluation metrics, then a repeated measures Analysis of Variance (ANOVA) procedure is conducted [76]. The statistical testing is done using the Tukey multiple pairwise testing procedure for all pairwise comparisons [76]. The Tukey test is a parametric test. The output of the Tukey test is a confidence limit for each pair of differences $v_{M_i} - v_{M_j}$. If the confidence limit includes zero, then statistically $v_{M_i} = v_{M_j}$ for the particular pair of models ($M_i$ and $M_j$) being compared. If the entire confidence limit is less than zero, then statistically $v_{M_i} < v_{M_j}$ for the particular pair of models being compared. If the entire confidence limit is greater than zero, then statistically $v_{M_i} > v_{M_j}$ for the particular pair of models being compared. These confidence limits can thus be used to rank all the models in terms of the evaluation metric under consideration. When assumption of a Gaussian parametric distribution is valid, using a parametric statistical test is superior to using a nonparametric test.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Chapter 5

# Airborne Fuel Burn

In this chapter, the statistical techniques described in Chapter 4 are applied to FDR data in order to build fuel flow rate and fuel burn models for the airborne phases of flight. Separate models are developed for each aircraft/engine type and for each of the airborne phases (that is, ascent, cruise, and descent). Additionally, separate models are built to analyze the fuel flow rate and fuel burn in climb out and approach, motivated by the observation that these phases occur at low altitudes, close to the surface [35].

This chapter begins by explaining how a physical understanding of aircraft and engine behavior has been used to identify features important for fuel flow rate prediction. It then proceeds to describe two fuel flow rate prediction algorithms. The proposed statistical models are evaluated using the metrics developed in Chapter 4. An important aspect of this chapter is the comparison of model predictive performance to that of currently used APMs such as the ICAO Databank, BADA Family 3, and the SFI models. In this manner, we demonstrate the improvements in performance afforded by the proposed modeling methodologies.

## 5.1 Candidate model features

The features of a model refer to the variables (and their transformations) that are included in the model. Instead of a purely data-driven feature identification procedure, this thesis leverages a physical understanding of aircraft dynamics to engineer the appropriate features for fuel flow

rate modeling.

Four principal forces act on an aircraft in flight, namely, the net thrust generated by the engines which propels the aircraft forward, the lift generated (primarily by the wings) which keeps it in the air, the drag or the resistance offered by the air, and the weight of the aircraft (Figure 5-1). Considering the aircraft as a point mass, the equations of motion can be written, in a simplified form, as follows:

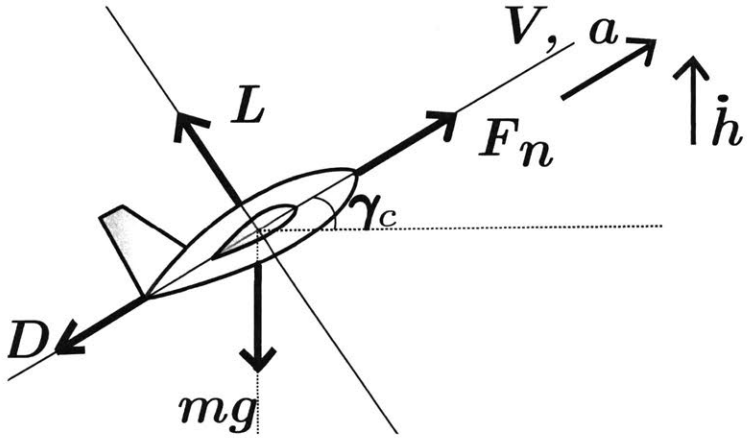



Figure 5-1: Simplified free-body diagram showing the forces acting on an aircraft in flight.

$$L = mg\cos\gamma_c \tag{5.1}$$

$$F_n = D + mg\sin\gamma_c + ma \tag{5.2}$$

$$L = q\mathcal{S}C_L \tag{5.3}$$

$$\sin\gamma_c = \frac{\dot{h}}{V} \tag{5.4}$$

$$D = q\mathcal{S}(C_{D_0} + C_{D_2}C_L^2) \tag{5.5}$$

$$q = \frac{1}{2}\rho_\infty V^2 \tag{5.6}$$

$$a = \frac{dV}{dt} \tag{5.7}$$

Here, $L$ is the lift on the aircraft, $m$ is the aircraft gross mass, $g$ is the acceleration due to gravity, $\gamma_c$ is the flight path angle, $F_n$ is the aircraft net thrust from all the engines, $D$ is the aircraft drag, $a$ is the aircraft acceleration in the net thrust direction, $q$ is the dynamic pressure,

$\mathcal{S}$ is the reference wing area, $C_L$ is the aircraft lift coefficient, $\dot{h}$ is the vertical speed, $V$ is the true airspeed, $C_{D_0}$ and $C_{D_2}$ are aircraft drag coefficients, $\rho_\infty$ is the ambient air density, and $t$ is the time. In this simplified dynamics model of the aircraft, the aircraft angle of attack and the bank angle are neglected. These quantities are small for large portions of the flight. More importantly, this simplified model is commonly employed in literature and is not found to introduce a significant error in the fuel burn prediction as compared to a more sophisticated dynamics model [15, 18]. Combining Equations 5.1–5.7, we have

$$F_n = q\mathcal{S}C_{D_0} + \frac{C_{D_2}m^2g^2}{q\mathcal{S}} - \frac{C_{D_2}m^2g^2\dot{h}^2}{q\mathcal{S}V^2} + mg\frac{\dot{h}}{V} + m\frac{dV}{dt}. \tag{5.8}$$

The average fuel flow rate per engine ($\dot{m}_f$) can be related to the net thrust via the Thrust Specific Fuel Consumption (TSFC) as

$$\dot{m}_f = \frac{\text{TSFC} \times F_n}{N_{\text{eng}}} \tag{5.9}$$

where, $N_{\text{eng}}$ is the number of engines. For a particular engine type, the corrected TSFC can be assumed to be a function of the aircraft Mach number, the corrected engine parameters (such as corrected net thrust for the engine), and the engine component efficiencies [56, 39], given as

$$\frac{\text{TSFC}}{\sqrt{\theta_\infty}} = f_{\text{TSFC}}\left(M_\infty, \frac{F_{n,\text{pe}}}{\delta_\infty}, \nu_{\text{eng}}\right). \tag{5.10}$$

Here, $M_\infty = \frac{V}{\sqrt{\gamma RT_\infty}}$ is the aircraft Mach number, $\delta_\infty$ is the ambient pressure ($P_\infty$) divided by the International Standard Atmosphere (ISA) sea level static pressure (= 101,325 Pa), $\theta_\infty$ is the ambient temperature ($T_\infty$) divided by the ISA sea level static temperature (= 288.15 K), $F_{n,\text{pe}}$ is the net thrust per engine $\left(F_{n,\text{pe}} = \frac{F_n}{N_{\text{eng}}}\right.$, assuming all the engines on the aircraft are producing equal net thrust$\left.\right)$, and $\nu_{\text{eng}}$ represents the engine component efficiencies. The parameter $\gamma$ is the adiabatic constant for air (= 1.4) and $R$ is the gas constant for air (= 287.05 J/kg/K). The above equations and relations (Equations 5.8–5.10) reveal the following functional dependency for the averaged per engine fuel flow rate for a particular aircraft/engine type (neglecting constants for

the particular aircraft/engine type):

$$\dot{m}_f\Big|_{\text{aircraft/engine type}} = f_{\dot{m}_f}\left(q\mathcal{S}, m, \frac{\dot{h}}{V}, V, \frac{dV}{dt}, \sqrt{T_\infty}, C_{D_0}, C_{D_2}, \nu_{\text{eng}}\right) \tag{5.11}$$

The aim of this thesis is to develop *open* models that map the aircraft trajectory to fuel flow rate profiles, which can then be used to estimate the fuel flow rate for a flight, given its trajectory from a surveillance system. Trajectory data from surveillance systems typically comprise the aircraft latitude, longitude, altitude, ground speed, climb rate, and course angle, as functions of time and are easily accessible. Many of the factors that affect fuel flow rate (Equation 5.11), such as the ambient conditions, aircraft mass, true airspeed, the drag coefficients, and the engine component efficiencies cannot be obtained from just the surveillance systems. A model that incorporates these features would need access to high-fidelity datasets beyond just surveillance data (for example, flight recorder archives, airline or manufacturer data, etc.), not all of which are easily accessible. Therefore, reasonable modeling assumptions need to be made in order to restrict the feature set, to the extent possible, to features which are extractable from trajectory data from surveillance systems.

### 5.1.1 Modeling assumptions

- Knowledge of the values of the ambient temperature and density at different locations along the flight trajectory needs access to high-fidelity weather data. When the true ambient conditions are not available, a surrogate density value is assumed according to the International Standard Atmosphere (ISA) model [103], which models the density as a function of the altitude (available from ground-based surveillance systems). Since the altitude alone determines the density in the ISA model, it is sufficient to retain only one ambient parameter (such as the density) in the model, as knowledge of the density gives the temperature too via the altitude. The equations for the density variation in the ISA

model are

$$\rho_\infty = \begin{cases} \rho_{\mathrm{SL}}\left(1+\beta_{\mathrm{ISA}}\frac{h}{T_{\mathrm{SL}}}\right)^{-\frac{g}{\beta_{\mathrm{ISA}}R}-1}, & \text{if } h < 11,000 \text{ m, and} \\ \rho_{11}\exp\left(\frac{-g}{RT_{11}}(h-h_{11})\right), & \text{if } 11,000 \text{ m} \leq h < 20,000 \text{ m.} \end{cases} \tag{5.12}$$

Here, $\rho_{\mathrm{SL}}$ is the air density at sea level (= 1.225 kg·m$^{-3}$), $\beta_{\mathrm{ISA}}$ is the tropospheric lapse rate (= -0.0065 K·m$^{-1}$), $h$ is the altitude above mean sea level, $T_{\mathrm{SL}}$ is the air temperature at sea level (= 288.15 K), $g$ is the acceleration due to gravity (= 9.81 m·s$^{-2}$), $R$ is the gas constant for air (= 287.05 J·kg$^{-1}$K$^{-1}$), $\rho_{11}$ is the air density 11,000 m above sea level (= 0.364 kg·m$^{-3}$), $T_{11}$ is the air temperature 11,000 m above sea level (= 216.65 K) and $h_{11}$ is the height of the tropopause start above mean sea level (= 11,000 m).

- Aircraft performance depends on the true airspeed $(V)$, which is the aircraft velocity with respect to the surrounding air. Due to the effect of wind, this speed differs from the ground speed, namely, the speed relative to the ground. The trajectory (or ground track) of an aircraft, as observed by a surveillance system, only provides information on the ground speed. The calculation of the true airspeed will require additional wind data of the same spatial and temporal resolution as the trajectory information. Historical wind data are available at a much coarser resolution (for example, airport surface winds are archived at 30 or 60 minute resolutions [104]; winds aloft are nowcast on a 13 km grid at hourly intervals at 51 discrete altitudes [105]; upper air wind and temperature data are also only recorded at a 1.25° or 138 km resolution globally [106]). As the value of the true airspeed cannot be derived from the trajectory with the same fidelity as the other variables, we do not include it as an input feature. Instead, we consider the ground speed $\left(V_{\mathrm{GS}}\right)$, which can be derived from trajectory data. We expect that the neglect of the effects of wind will be reflected in larger variability, and consequently, larger prediction errors as well as prediction intervals.

- The rate of change of the true airspeed $\left(\frac{dV}{dt}\right)$ (as required in Equation 5.11), is also not observed. The numerical derivative of the ground speed $\left(\frac{\Delta V_{\mathrm{GS}}}{\Delta t}\right)$ is therefore used as a feature. The values of the numerical derivative are smoothed through a low pass filter

before using them for analysis to remove noise arising from numerical differentiation.

- The Base of Aircraft Data (BADA) assumptions on the drag coefficients are assumed to hold. In ascent and cruise, the drag coefficients are considered to be aircraft-type specific constants. In descent, the drag coefficients are assumed to have discrete levels depending on the altitude of the aircraft with respect to the mean sea level elevation of the arrival airport ($h_{\mathrm{ATD}}$), the aircraft speed, and the aircraft gross mass. This dependence results from the different configurations that the aircraft adopts as it comes in to land (due to deployment of flaps, slats, landing gear), which change the aircraft drag coefficients.

- The engine component efficiencies depend on factors like the engine operating point, the engine age, and the level of maintenance undergone by the engine. Because trajectory data alone do not give information on the engine component efficiencies or factors influencing the efficienies, $\nu_{\mathrm{eng}}$ is not included as a predictor variable.

### 5.1.2 Feature selection

Based on the assumptions in Section 5.1.1, the functional dependence of the averaged per engine fuel flow rate for a particular aircraft/engine type in Equation 5.11 can be approximately re-written as in Equation 5.13:

$$\dot{m}_f\Big|_{\text{aircraft/engine type}} \approx \begin{cases} \dot{m}_{f_{(\mathrm{as})}}\left(q_{\mathrm{GS}}S, m, \frac{\dot{h}}{V_{\mathrm{GS}}}, V_{\mathrm{GS}}, \frac{\Delta V_{\mathrm{GS}}}{\Delta t}\right), & \text{in ascent,} \\ \dot{m}_{f_{(\mathrm{co})}}\left(q_{\mathrm{GS}}S, m, \frac{\dot{h}}{V_{\mathrm{GS}}}, V_{\mathrm{GS}}, \frac{\Delta V_{\mathrm{GS}}}{\Delta t}\right), & \text{in climb out,} \\ \dot{m}_{f_{(\mathrm{cr})}}\left(q_{\mathrm{GS}}S, m, \frac{\dot{h}}{V_{\mathrm{GS}}}, V_{\mathrm{GS}}, \frac{\Delta V_{\mathrm{GS}}}{\Delta t}\right), & \text{in cruise,} \\ \dot{m}_{f_{(\mathrm{de})}}\left(q_{\mathrm{GS}}S, m, \frac{\dot{h}}{V_{\mathrm{GS}}}, V_{\mathrm{GS}}, \frac{\Delta V_{\mathrm{GS}}}{\Delta t}, h_{\mathrm{ATD}}\right), & \text{in descent, and} \\ \dot{m}_{f_{(\mathrm{ap})}}\left(q_{\mathrm{GS}}S, m, \frac{\dot{h}}{V_{\mathrm{GS}}}, V_{\mathrm{GS}}, \frac{\Delta V_{\mathrm{GS}}}{\Delta t}, h_{\mathrm{ATD}}\right), & \text{in approach.} \end{cases} \tag{5.13}$$

In Equation 5.13, the subscripts $(\mathrm{as})$, $(\mathrm{co})$, $(\mathrm{cr})$, $(\mathrm{de})$, and $(\mathrm{ap})$ refer to the full ascent, only climb out, full cruise, full descent, and only approach phases, respectively. The different subscripts indicate that for each aircraft/engine type, different models are built for each of these five phases. $q_{\mathrm{GS}} = \frac{1}{2}\rho_{\infty}V_{\mathrm{GS}}^2$ is the dynamic pressure based on the ground speed (with ambient density given by the ISA model if needed).

Thus, in the regression models to be built, the input/independent/predictor variables are the following:

- Ground speed-based dynamic pressure multiplied by the reference wing area ($q_{\mathrm{GS}}S$, in N)
- Aircraft gross mass ($m$, in kg)
- Ratio of the aircraft vertical speed to its ground speed ($\frac{\dot{h}}{V_{\mathrm{GS}}}$, a dimensionless quantity)
- Aircraft ground speed ($V_{\mathrm{GS}}$, in $\mathrm{m{\cdot}s^{-1}}$)
- Rate of change of the aircraft ground speed ($\frac{\Delta V_{\mathrm{GS}}}{\Delta t}$, in $\mathrm{m{\cdot}s^{-2}}$)
- Aircraft altitude above the mean sea level elevation of the arrival airport ($h_{\mathrm{ATD}}$, in m) (additional feature only in descent and approach)

In all the phases, the output/dependent/predicted variable is the averaged fuel flow rate per engine ($\dot{m}_f$, in $\mathrm{kg{\cdot}s^{-1}}$). It should be noted that except for the aircraft gross mass, all other input variables can be obtained from surveillance systems.

The predictor variables chosen are the primary factors affecting the fuel flow rate. However, many other factors (for example, aging, component deterioration, engine bleed, power offtake, aircraft attitudes and attitude rates, wind vector, etc.) also affect the fuel flow rate at the same values of the primary factors. These secondary factors are difficult to model due to their absence from trajectory data from surveillance systems. However, it is assumed that the prediction intervals given by the statistical models can give a cumulative extent of the uncertainty in fuel flow rate resulting from these unmodeled secondary factors.

## 5.2 Model training

As explained in Section 4.1.1, the FDR dataset for each aircraft/engine type in each phase is divided into three smaller datasets. 65% of the flights are randomly chosen to constitute the training set, 15% of the flights are randomly chosen to constitute the validation set, and the remaining flights constitute the test set. Each observation (data point) in the training, validation, and test sets corresponds to one FDR instantaneous recording. The training set, consisting of the ground truth values of all the input and the output features, is used for training the model. Model training involves building the statistical model which can map the predictor variables to the output fuel flow rate. The validation (or development) set is used for model validation, where the different models trained using different regression techniques are evaluated for their predictive performance in order to choose the "best" model. The test set is used to evaluate and report the predictive performance of the chosen model. It is important to note that the three sets are mutually exclusive and all the observations of a particular flight are present in only one of the three sets. Table 5.1 gives an idea of the size of the different sets by enumerating the number of flights and the number of observations in each phase for each aircraft type.

Prior to training, all the variables are standardized, that is shifted by the sample mean and then scaled by the sample standard deviation of the respective variables in the training datasets. Ordinary Least Squares (OLS) regression (with quadratic basis functions and Gaussian noise), Classification and Regression Trees (CART), regression tree-based Least Squares Boosting (LSB), and Gaussian Process Regression (GPR) with different kernel functions are applied to the standardized variables in the training sets to build the models. In GPR, the hyperparameters, being all positive, are given a broad gamma prior with mode 1 and variance 100 (for lack of specific prior knowledge). Exact GP is used for inferring the hyperparameters in cruise. In ascent, climb out, descent, and approach, due to the large size of the training datasets, GP with the Fully Independent Conditional (FIC) approximation is used for hyperparameter inference. For this FIC approximation, 150 observations are randomly chosen from the appropriate training set (depending on the phase) to form the inducing inputs and variables (the number of inducing variables is chosen by doing a study on the validation dataset of how the prediction error and the testing time vary with the number of inducing variables). It is important to note

Table 5.1: Details of training, validation, and test datasets drawn from the FDR-I dataset, for each aircraft type, and in each airborne phase.

| A/C Type | Phase | Training Set | | Validation Set | | Test Set | |
|---|---|---|---|---|---|---|---|
| | | # Flights | # Obs. | # Flights | # Obs. | # Flights | # Obs. |
| A319-112 | Ascent | 85 | 17,810 | 19 | 4,016 | 26 | 5,609 |
| | Climb out | | 5,469 | | 1,197 | | 1,636 |
| | Cruise | | 1,146 | | 205 | | 303 |
| | Descent | | 42,398 | | 8,984 | | 12,812 |
| | Approach | | 20,609 | | 4,504 | | 6,890 |
| A320-214 | Ascent | 110 | 26,101 | 25 | 6,260 | 34 | 8,256 |
| | Climb out | | 5,855 | | 1,340 | | 1,810 |
| | Cruise | | 1,416 | | 308 | | 481 |
| | Descent | | 54,350 | | 12,409 | | 16,592 |
| | Approach | | 24,562 | | 5,560 | | 7,411 |
| A321-111 | Ascent | 76 | 18,261 | 18 | 4,241 | 23 | 6,171 |
| | Climb out | | 7,313 | | 1,712 | | 2,279 |
| | Cruise | | 933 | | 190 | | 307 |
| | Descent | | 34,110 | | 8,184 | | 11,673 |
| | Approach | | 16,310 | | 4,034 | | 5,006 |
| A330-202 | Ascent | 55 | 15,825 | 12 | 3,712 | 17 | 5,140 |
| | Climb out | | 5,551 | | 1,325 | | 1,807 |
| | Cruise | | 3,443 | | 672 | | 1,023 |
| | Descent | | 27,224 | | 5,805 | | 8,258 |
| | Approach | | 16,885 | | 3,502 | | 5,127 |
| A330-243 | Ascent | 65 | 16,011 | 15 | 3,985 | 20 | 5,485 |
| | Climb out | | 5,232 | | 1,167 | | 1,652 |
| | Cruise | | 2,317 | | 583 | | 908 |
| | Descent | | 32,445 | | 7,830 | | 10,228 |
| | Approach | | 20,660 | | 4,613 | | 5,945 |
| A340-541 | Ascent | 34 | 9,543 | 8 | 1,998 | 10 | 2,702 |
| | Climb out | | 4,464 | | 955 | | 1,222 |
| | Cruise | | 1,736 | | 390 | | 482 |
| | Descent | | 17,438 | | 4,238 | | 5,187 |
| | Approach | | 9,892 | | 2,574 | | 3,003 |
| B767-300 | Ascent | 59 | 11,378 | 14 | 2,608 | 18 | 3,544 |
| | Climb out | | 4,789 | | 1,084 | | 1,375 |
| | Cruise | | 2,683 | | 406 | | 944 |
| | Descent | | 28,197 | | 6,388 | | 8,294 |
| | Approach | | 16,784 | | 3,918 | | 5,197 |
| B777-300ER | Ascent | 85 | 19,338 | 20 | 4,481 | 26 | 5,748 |
| | Climb out | | 5,474 | | 1,334 | | 1,706 |
| | Cruise | | 4,749 | | 1,198 | | 1,493 |
| | Descent | | 39,272 | | 9,475 | | 12,167 |
| | Approach | | 20,756 | | 5,138 | | 6,827 |

that for each aircraft type, different models are trained for the different phases of flight. Thus, the models developed are locally expert models, where each model is intended to predict the fuel flow rates only in a particular phase of flight. The combined predictive performance of

these locally expert models is expected to be better than that of a single model trained to predict the fuel flow rates for the complete airborne trajectory. Once the models are trained, they can be used to determine the point estimates, the prediction intervals, and the predictive distributions of the fuel flow rate at a new input vector.

## 5.3 Fuel flow rate prediction

As mentioned in Section 5.1.2, aircraft gross mass is the only input feature which cannot be directly obtained from trajectory data. Depending on how the gross mass is handled as a predictor, two different algorithms have been developed for fuel flow rate prediction.

### 5.3.1 Batch prediction algorithm

Aircraft gross mass decreases with time as fuel is consumed. In general, the exact values of the instantaneous aircraft mass for any particular flight cannot be obtained from its trajectory data or any other easily available dataset. To partially overcome this problem, in this section, the takeoff weight (TOW, $m_{\mathrm{TO}}$) is used as a surrogate for the actual instantaneous aircraft mass during flight. It should be noted that the TOW is also not available in general and would need to be estimated. However, a single value of the mass at takeoff is easier to obtain/estimate as compared to estimating the entire instantaneous aircraft mass profile during flight. Moreover, the use of TOW as a predictor variable enables simultaneous prediction of the fuel flow rates at all the points in a particular phase of a flight (hence the name, 'batch' prediction algorithm). Therefore, for the analysis in this section, the predictor features are the trajectory-based variables described in Section 5.1 along with the takeoff weight (TOW). The predicted variable is the averaged fuel flow rate per engine. The ground truth values of the input and the output variables (including the ground truth values of the TOW and the ambient air density) from the FDR training datasets are used for training the models using the different regression techniques mentioned in Section 5.2.

In order to determine model generalized predictive performance on a new flight trajectory, it needs to be evaluated on an unseen prediction set not used for training. While evaluating the

predictive performance of the trained models on an unseen prediction set (such as the validation or the test sets), the values of the ambient air density as calculated by using the ISA approximation (Equation 5.12) are used (as new flight trajectory data will not have the actual density values). The values of the other input features (including the TOW) for an unseen observation in the validation or test datasets are assumed to be known from the FDR data for the analysis done in this chapter. However, in order to get a complete estimate of model generalized predictive performance, the values of these input features should, in reality, be obtained from surveillance data as FDR data for a new flight may not be available. More importantly, TOW also needs to be estimated since its value for a new flight will, in general, be unknown. In the analysis done in this chapter, FDR values of the input variables for new flights are assumed to be known so as to show model predictive performance on unseen data without incorporating errors arising due to differences in the FDR- and surveillance data-reported values of the input variables. Model predictive performance on unseen data when the TOW is unknown and also needs to be estimated will be described in a later chapter (Chapter 7).

After models are trained using different regression techniques, the evaluation metrics described in Section 4.5 are calculated for the different models on the respective validation sets. MAE, PC, PLL for the different models are compared among one another using statistical multiple comparison tests at an at most 5% family significance level (Section 4.6) to find the "best" statistical model (in terms of the accuracy of the point estimate, the uncertainty estimate, and the estimate of the predictive distribution) for each aircraft type in each phase. Since tree-based models (CART and LSB) are nonprobabilistic in nature, their prediction intervals are determined using bootstrapping (Section 4.3.3). The PLL values for OLS regression models are calculated with respect to a Student's t-distribution (Equations 4.11–4.14, 4.40). PLL values cannot be determined for the tree-based models as they do not give a predictive distribution. The PLL values for GPR models are calculated with respect to a Gaussian distribution (Equation 4.39). Table 5.2 shows the values of the MAE, PC, PLL on the validation datasets which are used for model selection. The table also shows the "best" model finally chosen for each aircraft type in each phase. Due to lack of space, the table does not show values of the prediction coverage for the tree-based models (CART and LSB). However, both the tree-based methods perform

poorly in terms of the prediction coverage. For example, for the A321-100, the mean PC across the different flights in the validation set ranges from 43.9% to 57.3% for the CART models and from 40.7% to 42.9% for the LSB models. Similar low values of PC are also observed for the other aircraft types.

From Table 5.2, it can be observed that the GPR models give, statistically significantly, the best predictive performance on validation data for all the aircraft types in all the phases, when collectively evaluated on the different evaluation metrics. Thus, GPR models are selected as the final models for fuel flow rate prediction using the batch prediction algorithm. Different kernel functions are seen to give the best performance for different aircraft types and phases. It should be noted that though the GPR models take the longest time to be trained as compared to the other models, they give fast predictions (on the order of a few seconds). Hence, once these models are trained, the time taken to predict the fuel flow rate for a new flight is small.

### 5.3.2 One-step prediction algorithm

In this algorithm, the instantaneous aircraft gross mass (and not the TOW) is used as a predictor variable (along with the other trajectory-based variables). Aircraft gross mass decreases with time as fuel is consumed. While training, the actual ground truth values of the instantaneous aircraft gross mass and the other predictor variables from the FDR training datasets are used as inputs. However, the main aim of model building in this thesis is to use the models for predicting the fuel flow rates for new, unseen flights not used for training. For a new flight, the actual value of the instantaneous aircraft gross mass may not be available (since FDR data for the new flight could be unavailable and surveillance data cannot record the mass of the aircraft as it flies through the air). Hence, for a new flight, the value of the aircraft gross mass at a particular time instant needs to be estimated in order for it to be used as an input to further estimate the fuel flow rate at that time instant.

In this section, an algorithm is developed which estimates both the instantaneous aircraft gross mass and the fuel flow rate for the airborne phases of a flight. This algorithm operates on the knowledge that the reduction in gross mass with time is due to fuel burn only. For this algorithm to work, it is assumed that the values of the aircraft gross mass and the fuel flow

Table 5.2: Fuel flow rate predictive performance of statistical models (using TOW as a predictor) on validation data. Each entry shows the mean of the evaluation metric across all the flights in the validation dataset. Except for the Predictive Log Likelihood (PLL), all the evaluation metrics are calculated on de-standardized data (that is, data at their original location and scale and not shifted by the mean and scaled by the standard deviation of the variables in the training dataset). 'GPR1', 'GPR2' 'GPR3', 'GPR4' refer to the GPR models with the DPSE, DPE, DPM3/2, and DPM5/2 kernels, respectively.

| A/C Type | Phase | MAE | | | | | | | PC | | | | | PLL | | | | | Best |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *OLS* | *CART* | *LSB* | *GPR1* | *GPR2* | *GPR3* | *GPR4* | *OLS* | *GPR1* | *GPR2* | *GPR3* | *GPR4* | *OLS* | *GPR1* | *GPR2* | *GPR3* | *GPR4* | Model |
| A319-112 | Ascent | 5.3 | 6.1 | 5.5 | 4.9 | 5.1 | 4.9 | 4.9 | 94.8 | 93.7 | 95.9 | 93.7 | 93.5 | -332.1 | -46.4 | -62.3 | -47.0 | -57.8 | **GPR2** |
| | Climb out | 3.1 | 4.5 | 4.0 | 3.3 | 3.1 | 3.3 | 3.3 | 87.4 | 84.8 | 88.9 | 83.9 | 84.2 | -110.1 | -74.5 | -66.2 | -76.2 | -77.0 | **GPR2** |
| | Cruise | 13.7 | 14.6 | 13.1 | 13.7 | 12.6 | 12.6 | 12.5 | 94.9 | 91.4 | 92.4 | 95.0 | 96.0 | -15.1 | -11.1 | -10.5 | -9.7 | -9.6 | **GPR4** |
| | Descent | 24.5 | 23.8 | 22.3 | 20.2 | 20.0 | 20.1 | 20.1 | 94.5 | 94.6 | 94.8 | 94.4 | 94.2 | -674.3 | -392.3 | -399.4 | -397.6 | -399.4 | **GPR2** |
| | Approach | 16.7 | 20.0 | 18.4 | 16.1 | 15.8 | 15.9 | 16.1 | 95.0 | 94.0 | 95.9 | 94.6 | 94.1 | -336.4 | -196.6 | -190.3 | -192.3 | -194.9 | **GPR2** |
| A320-214 | Ascent | 5.8 | 6.2 | 5.6 | 4.6 | 5.3 | 4.6 | 4.5 | 93.4 | 94.7 | 94.4 | 94.9 | 94.7 | -418.4 | -19.8 | -91.4 | -22.8 | -20.1 | **GPR1** |
| | Climb out | 3.4 | 4.7 | 4.1 | 3.7 | 3.5 | 3.4 | 3.4 | 94.2 | 90.7 | 95.8 | 92.4 | 92.2 | -95.9 | -48.0 | -42.7 | -41.2 | -42.0 | **GPR2** |
| | Cruise | 11.6 | 14.0 | 12.4 | 12.8 | 13.6 | 12.1 | 13.1 | 90.7 | 91.6 | 95.4 | 94.0 | 91.5 | -19.1 | -17.3 | -14.7 | -14.2 | -18.0 | **GPR2** |
| | Descent | 26.8 | 24.9 | 23.8 | 24.3 | 24.6 | 23.6 | 24.0 | 93.1 | 91.9 | 92.1 | 92.2 | 92.6 | -746.7 | -443.1 | -457.3 | -441.5 | -438.2 | **GPR4** |
| | Approach | 17.9 | 18.9 | 16.4 | 16.4 | 16.6 | 15.8 | 16.1 | 94.0 | 94.1 | 95.6 | 94.2 | 94.1 | -319.6 | -166.3 | -167.4 | -165.8 | -167.8 | **GPR2** |
| A321-111 | Ascent | 4.1 | 4.4 | 3.9 | 3.4 | 4.0 | 3.4 | 3.3 | 95.8 | 93.6 | 95.5 | 94.2 | 94.1 | -302.8 | -15.6 | -37.9 | -13.9 | -9.5 | **GPR1** |
| | Climb out | 4.3 | 5.3 | 4.6 | 4.2 | 4.2 | 3.7 | 3.8 | 95.8 | 89.7 | 92.8 | 90.8 | 91.1 | -104.3 | -43.4 | -39.1 | -36.4 | -36.3 | **GPR2** |
| | Cruise | 15.9 | 18.2 | 14.4 | 14.6 | 15.0 | 15.0 | 14.8 | 97.8 | 95.0 | 94.4 | 94.0 | 94.5 | -13.1 | -8.3 | -8.5 | -8.7 | -8.6 | **GPR1** |
| | Descent | 25.0 | 25.1 | 23.0 | 22.3 | 22.1 | 22.1 | 22.7 | 95.2 | 93.0 | 95.1 | 92.7 | 92.8 | -644.2 | -387.1 | -380.4 | -388.7 | -393.6 | **GPR2** |
| | Approach | 18.7 | 18.0 | 17.7 | 17.3 | 17.3 | 17.4 | 17.4 | 94.1 | 92.8 | 93.2 | 92.7 | 92.9 | -333.6 | -194.0 | -188.8 | -194.3 | -191.8 | **GPR2** |
| A330-202 | Ascent | 4.3 | 4.4 | 3.9 | 3.9 | 4.0 | 3.9 | 3.9 | 95.1 | 93.5 | 96.5 | 93.6 | 93.5 | -435.2 | -18.8 | 9.3 | -25.3 | -27.2 | **GPR2** |
| | Climb out | 4.8 | 3.8 | 3.6 | 5.1 | 3.9 | 4.8 | 4.8 | 92.5 | 86.2 | 92.4 | 86.9 | 85.9 | -166.8 | -105.6 | -67.5 | -95.1 | -96.5 | **GPR2** |
| | Cruise | 10.1 | 7.8 | 7.4 | 7.6 | 9.0 | 7.5 | 7.5 | 92.1 | 98.9 | 99.2 | 99.6 | 99.6 | -84.1 | -51.2 | -56.9 | -54.0 | -52.0 | **GPR2** |
| | Descent | 34.4 | 29.7 | 28.0 | 27.1 | 27.0 | 26.1 | 28.6 | 93.7 | 92.0 | 92.3 | 92.5 | 91.6 | -701.8 | -413.9 | -412.4 | -405.3 | -439.6 | **GPR3** |
| | Approach | 29.6 | 26.7 | 24.1 | 23.4 | 25.9 | 23.7 | 23.6 | 92.1 | 91.3 | 92.9 | 91.1 | 91.2 | -450.1 | -265.9 | -271.4 | -268.6 | -267.1 | **GPR1** |
| A330-243 | Ascent | 3.8 | 4.6 | 3.9 | 3.1 | 3.6 | 3.2 | 3.1 | 97.7 | 97.6 | 97.4 | 97.5 | 97.5 | -333.6 | 52.5 | -10.4 | -47.8 | 55.7 | **GPR1** |
| | Climb out | 3.0 | 4.0 | 3.2 | 6.0 | 3.1 | 5.1 | 5.6 | 94.7 | 61.7 | 92.1 | 70.8 | 63.7 | -128.8 | -213.7 | -86.4 | -155.7 | -182.8 | **GPR2** |
| | Cruise | 12.2 | 10.3 | 8.9 | 11.1 | 12.1 | 10.9 | 10.8 | 95.4 | 97.1 | 93.3 | 97.2 | 97.5 | -50.7 | -35.0 | -31.0 | -31.2 | -32.3 | **GPR3** |
| | Descent | 30.6 | 20.4 | 20.6 | 25.6 | 22.7 | 23.4 | 25.5 | 92.3 | 89.0 | 92.4 | 90.5 | 89.3 | -834.0 | -591.5 | -468.3 | -510.8 | -582.5 | **GPR2** |
| | Approach | 27.4 | 22.0 | 21.1 | 23.3 | 22.2 | 23.4 | 23.4 | 91.0 | 88.4 | 90.3 | 88.7 | 88.5 | -527.1 | -305.1 | -284.2 | -306.2 | -304.6 | **GPR2** |
| A340-541 | Ascent | 4.7 | 5.1 | 4.6 | 5.0 | 4.5 | 4.0 | 4.1 | 94.9 | 89.9 | 94.0 | 95.6 | 95.7 | -344.1 | -89.8 | -59.3 | -42.1 | -46.9 | **GPR3** |
| | Climb out | 3.3 | 4.8 | 4.0 | 5.9 | 3.3 | 3.7 | 4.3 | 88.6 | 75.2 | 90.2 | 75.5 | 74.4 | -208.7 | -433.5 | -75.1 | -297.3 | -387.3 | **GPR2** |
| | Cruise | 12.5 | 8.7 | 10.8 | 12.9 | 17.3 | 15.7 | 14.1 | 79.7 | 83.5 | 79.5 | 86.9 | 85.2 | -236.9 | -94.4 | -108.0 | -84.7 | -91.2 | **GPR3** |
| | Descent | 25.2 | 25.7 | 23.5 | 20.5 | 21.9 | 21.2 | 21.3 | 95.2 | 93.8 | 95.7 | 94.1 | 93.8 | -738.7 | -430.2 | -444.0 | -435.3 | -438.1 | **GPR1** |
| | Approach | 21.6 | 24.8 | 21.5 | 21.5 | 18.9 | 20.8 | 21.0 | 93.8 | 90.1 | 94.7 | 91.1 | 91.4 | -480.8 | -352.1 | -311.6 | -342.7 | -340.0 | **GPR2** |
| B767-300 | Ascent | 5.1 | 5.0 | 5.0 | 5.1 | 5.0 | 5.1 | 5.1 | 96.5 | 90.8 | 96.6 | 92.4 | 92.0 | -279.4 | -91.9 | -64.3 | -83.7 | -87.3 | **GPR2** |
| | Climb out | 4.5 | 3.0 | 2.9 | 4.7 | 3.4 | 4.0 | 4.3 | 87.1 | 67.3 | 96.3 | 80.8 | 77.5 | -155.0 | -116.9 | -41.3 | -73.9 | -82.8 | **GPR2** |
| | Cruise | 20.7 | 18.4 | 18.4 | 18.0 | 26.2 | 21.5 | 17.1 | 80.3 | 94.7 | 96.1 | 95.6 | 97.9 | -37.4 | -14.5 | -41.4 | -25.3 | -18.8 | **GPR4** |
| | Descent | 31.0 | 18.6 | 19.1 | 19.0 | 20.4 | 17.6 | 17.8 | 95.1 | 93.3 | 95.0 | 93.7 | 93.8 | -646.4 | -218.7 | -220.0 | -196.4 | -194.8 | **GPR2** |
| | Approach | 26.6 | 20.4 | 18.4 | 18.9 | 18.5 | 17.2 | 18.0 | 91.5 | 93.1 | 96.4 | 93.0 | 92.7 | -455.1 | -183.0 | -176.9 | -177.9 | -183.4 | **GPR2** |
| B777-300ER | Ascent | 8.1 | 7.8 | 6.9 | 6.8 | 7.6 | 6.6 | 6.7 | 93.8 | 93.5 | 95.8 | 94.1 | 93.9 | -324.9 | -46.0 | -64.0 | -33.3 | -36.1 | **GPR3** |
| | Climb out | 8.4 | 7.8 | 7.2 | 6.5 | 7.2 | 6.2 | 6.3 | 92.6 | 92.0 | 95.7 | 93.2 | 92.8 | -103.5 | -8.3 | -14.4 | -7.4 | -7.4 | **GPR3** |
| | Cruise | 11.5 | 12.2 | 11.1 | 11.1 | 11.3 | 10.9 | 10.8 | 87.9 | 92.4 | 88.1 | 91.4 | 91.4 | -161.7 | -84.1 | -99.6 | -75.9 | -78.5 | **GPR4** |
| | Descent | 32.6 | 16.2 | 16.1 | 18.5 | 16.7 | 17.0 | 17.2 | 89.6 | 91.7 | 91.9 | 91.7 | 91.9 | -745.0 | -281.8 | -231.1 | -256.1 | -262.6 | **GPR2** |
| | Approach | 21.1 | 18.9 | 17.5 | 15.8 | 16.4 | 15.0 | 14.9 | 91.0 | 90.6 | 93.8 | 91.0 | 90.9 | -428.3 | -159.1 | -145.5 | -151.1 | -153.2 | **GPR2** |

rate are known at takeoff, which just precedes the first instant in ascent (or climb out). In case it is not known, the fuel flow rate value at the takeoff point can be estimated using the TOW, the trajectory information at the takeoff point, and the fuel flow rate models developed for the ascent phase. The TOW is thus needed to be known and these takeoff values, therefore, serve as initial conditions to the algorithm. At any instant of time, the aircraft gross mass is estimated using the estimated values of the gross mass and the fuel flow rate at the previous instant of time using the following equation:

$$m_{i+1} = m_i - N_{\text{eng}} \dot{m}_{f_i} \Delta t_{i+1,i} \tag{5.14}$$

Here, $m_{i+1}$ and $m_i$ are the aircraft gross mass values at the $(i+1)^{\text{th}}$ and the $i^{\text{th}}$ time instants, respectively, $\dot{m}_{f_i}$ is the averaged fuel flow rate per engine at the $i^{\text{th}}$ time instant, and $\Delta t_{i+1,i}$ is the time interval between the $(i+1)^{\text{th}}$ and the $i^{\text{th}}$ time instants. The time interval between successive instants depends on the sampling frequency of the reported trajectory data for the new flight. The algorithm as such, does not need to assume an explicit value of the time step a priori.

The estimated gross mass at the current instant then serves as one of the inputs (the exact values of the other trajectory-based inputs at the current instant are assumed to be already known) to the fuel flow rate model trained (with the instantaneous aircraft mass as one of the predictors) for the particular phase (ascent/cruise/descent) in which the current time instant lies. This procedure continues till the end of descent. In this algorithm, uncertain estimates of fuel flow rate are used to estimate the gross mass and these uncertain estimates of gross mass then serve as inputs to again estimate the fuel flow rate. Thus, instantaneous gross mass and fuel flow rate are estimated in tandem in a stepwise fashion. It is important to correctly propagate these uncertainties from one time instant to the next. The regression problem in this case is hence, one having uncertainty in input values - a feature not encountered in common regression problems (in which the inputs are assumed to be free of any uncertainty). The uncertainty is in the gross mass input variable and not the other input variables whose exact values are assumed to be easily available. The approach described in [107] is suitably adopted to propagate the uncertainty in the input vector to the output. Let $\mathcal{D}$ be the training dataset and $\mathbf{x}^*$ be a new un-

certain input at which the prediction $y^*$ is desired. For correct propagation of uncertainty from input to output, we are interested in determining the probability of the output prediction given the training data, marginalized over all possible values of the uncertain input (Equation 5.15). By using a Monte Carlo approach, the integral in Equation 5.15 is approximated by an average (Equation 5.16).

$$p(y^*|\mathcal{D}) = \int p(y^*|\mathbf{x}^*, \mathcal{D}) p(\mathbf{x}^*) d\mathbf{x}^* \tag{5.15}$$

$$\approx \frac{1}{n_s} \sum_{j=1}^{n_s} p(y^*|\mathbf{x}_\mathbf{j}^*, \mathcal{D}) \tag{5.16}$$

where, $n_s$ is the number of samples and $\mathbf{x}_\mathbf{j}^*$ are samples of the uncertain input $\mathbf{x}^*$. For a GPR model, $p(y^*|\mathbf{x}_\mathbf{j}^*, \mathcal{D})$ is a Gaussian density. Thus, for a GPR model, $p(y^*|\mathcal{D})$ can be approximated by a Gaussian Mixture density with $n_s$ number of equally-weighted components. The approximation arises due to the use of a finite number of samples.

Algorithm 1 shows the one-step algorithm for fuel flow rate prediction. In the algorithm, subscripts 'prev' and 'curr' refer to the previous and the current time instants, respectively. 'IndexAscentStart' refers to the index of the point in the flight trajectory where ascent begins (the point immediately succeeding takeoff). 'IndexTouchdown' refers to the index of the point in the flight trajectory where the aircraft touches down at the arrival airport (just after the end of descent). $\mathbf{x}_{-\mathbf{m}}^*$ refers to the input vector where prediction is desired with the gross mass input variable removed. 'MeanPredGPR($\mathbf{x}^*$)' and 'VarPredGPR($\mathbf{x}^*$)' refer to the mean and the variance, respectively, of the prediction from the GPR fuel flow rate model (trained for the appropriate phase of flight, given in the subscripted parentheses) at a given input $\mathbf{x}^*$. The subscript '$1 : n_s$' refers to the vector formed by taking all $n_s$ samples together. In this study, 100 samples are used for the Monte Carlo integration approach described in Equation 5.16. The number of samples is chosen after considering the simulation time. $\mathcal{N}(\mu, \sigma^2)$ refers to a univariate Gaussian distribution with mean $\mu$ and variance $\sigma^2$. $\mathcal{GM}(\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \boldsymbol{w})$ refers to a Gaussian Mixture distribution (of univariate Gaussians) with vector of component means $\boldsymbol{\mu}$, vector of component variances $\boldsymbol{\sigma^2}$, and vector of component weights $\boldsymbol{w}$. The mean of the vector of fuel flow rate samples ($\dot{\mathbf{m}}_{\mathbf{f}_{\mathbf{curr}}}$), generated according to Algorithm 1, gives the average

predicted fuel flow rate per engine at a particular instant. The 95% prediction interval for the predicted fuel flow rate per engine at a particular instant is given by the inter-percentile range between the $2.5^{\text{th}}$ and the $97.5^{\text{th}}$ percentiles of the fuel flow rate samples ($\dot{\mathbf{m}}_{\mathbf{f}_{\mathbf{curr}}}$).

**Algorithm 1** Sequentially predicting aircraft gross mass and fuel flow rate in the airborne phases for a given flight (one-step prediction).

$m_{\text{prev}} \leftarrow \text{TakeOffGrossMass}$
$\dot{m}_{f_{\text{prev}}} \leftarrow \text{TakeoffFuelFlowRatePerEngine}$
$i \leftarrow \text{IndexAscentStart}$
$m_{\text{curr}} \leftarrow m_{\text{prev}} - N_{\text{eng}} \times \dot{m}_{f_{\text{prev}}} \times \Delta t_{\text{curr,prev}}$
$\mu_{\text{GPR}} \leftarrow \text{MeanPredGPR}_{(\text{as})}([\mathbf{x}^*_{-\mathbf{m}}, m]|_{\text{curr}})$
$\sigma^2_{\text{GPR}} \leftarrow \text{VarPredGPR}_{(\text{as})}([\mathbf{x}^*_{-\mathbf{m}}, m]|_{\text{curr}})$
**for** $j \leftarrow 1, 2, \ldots, n_s$ **do**
    $\dot{m}_{f_{\text{curr},j}} \sim \mathcal{N}(\mu_{\text{GPR}}, \sigma^2_{\text{GPR}})$
    $m_{\text{prev},j} \leftarrow m_{\text{curr}}$
    $\dot{m}_{f_{\text{prev},j}} \leftarrow \dot{m}_{f_{\text{curr},j}}$
$i \leftarrow i + 1$
**while** $i <$ IndexTouchdown **do**
    **for** $j \leftarrow 1, 2, \ldots, n_s$ **do**
        $m_{\text{curr},j} \leftarrow m_{\text{prev},j} - N_{\text{eng}} \times \dot{m}_{f_{\text{prev},j}} \times \Delta t_{\text{curr,prev}}$
        $\mu_{\text{GPR},j} \leftarrow \text{MeanPredGPR}_{(\text{as/cr/de as appropriate})}([\mathbf{x}^*_{-\mathbf{m}}, m_j]|_{\text{curr}})$
        $\sigma^2_{\text{GPR},j} \leftarrow \text{VarPredGPR}_{(\text{as/cr/de as appropriate})}([\mathbf{x}^*_{-\mathbf{m}}, m_j]|_{\text{curr}})$
    **for** $j \leftarrow 1, 2, \ldots, n_s$ **do**
        $\dot{m}_{f_{\text{curr},j}} \sim \mathcal{GM}(\boldsymbol{\mu}_{\boldsymbol{GPR},\mathbf{1:n_s}}, \boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{GPR},\mathbf{1:n_s}}, \mathbf{EqualComponentWeights})$
        $m_{\text{prev},j} \leftarrow m_{\text{curr},j}$
        $\dot{m}_{f_{\text{prev},j}} \leftarrow \dot{m}_{f_{\text{curr},j}}$
    $i \leftarrow i + 1$

Since this algorithm can predict fuel flow rate in a sequential fashion only (instead of predicting it simultaneously at all instants), it is named as the 'one-step' prediction algorithm. Table 5.3 compares the predictive performance of the one-step prediction algorithm with that of the batch prediction algorithm on the validation data for three aircraft types in climb out and approach. The kernel functions of the GPR models selected in Section 5.3.1 are used for both the algorithms. The algorithm comparison is done by using statistical tests to compare different evaluation metrics (Section 4.6). The null hypothesis for the statistical comparison is that an evaluation metric is equal for both the algorithms. The alternate hypothesis is that an evaluation metric on the two algorithms is unequal. Statistically significant p-values (at 5% significance

level) are highlighted in bold in Table 5.3. From Table 5.3, it is observed that in almost all the cases, batch prediction performs the same as or better than one-step prediction (in a statistically significant sense). Moreover, the one-step prediction algorithm is computationally slower than the batch prediction algorithm. Hence, the batch prediction algorithm is preferred over the one-step prediction algorithm.

Table 5.3: Comparison of the batch prediction (BP) and one-step prediction (OSP) algorithms on unseen data (not used for training) for three aircraft types in climb out and approach. Each cell shows the mean and the standard deviation (within parentheses) of the evaluation metric across all the flights. In order to maintain approximately the same number of flights for the different aircraft types, the flights are taken from the validation dataset for the A319-112 and the B777-300ER and from the combined validation and test datasets for the A340-541. The table also shows the p-values of the BP and OSP predictions differing on various metrics (that is, $MAE_{BP} \neq MAE_{OSP}$, $PC_{BP} \neq PC_{OSP}$, and $NLPI_{BP} \neq NLPI_{OSP}$).

| **A/C Type** | **Phase** | **MAE (%)** | | **PC (%)** | | **NLPI (%)** | | **Comparison: p-values** | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **BP** | **OSP** | **BP** | **OSP** | **BP** | **OSP** | **MAE** | **PC** | **NLPI** |
| A319-112 | Climb out | 3.1 (1.6) | 3.2 (1.6) | 88.9 (15.8) | 88.9 (15.0) | 14.1 (1.0) | 14.0 (1.1) | **0.016** | 0.856 | 0.171 |
| | Approach | 15.7 (4.3) | 15.9 (4.1) | 96.0 (4.4) | 95.3 (4.8) | 96.2 (14.2) | 97.6 (15.0) | 0.126 | **0.005** | **0.007** |
| A340-541 | Climb out | 4.1 (2.7) | 4.0 (2.7) | 83.7 (22.7) | 84.6 (22.1) | 14.4 (0.9) | 14.3 (0.8) | 0.528 | 0.384 | 0.145 |
| | Approach | 17.6 (3.4) | 17.8 (4.2) | 95.4 (3.7) | 96.3 (3.0) | 123.8 (20.0) | 129.6 (26.2) | 0.983 | 0.107 | 0.133 |
| B777-300ER | Climb out | 6.2 (1.7) | 6.5 (1.5) | 93.2 (4.3) | 91.8 (6.0) | 30.8 (4.5) | 30.9 (4.4) | **0.033** | 0.151 | 0.823 |
| | Approach | 15.6 (4.7) | 17.6 (6.4) | 94.9 (4.7) | 91.0 (6.4) | 98.8 (16.1) | 90.9 (18.1) | **0.015** | **2.5e-4** | **7.8e-4** |

The validation studies in Sections 5.3.1 and 5.3.2 result in GPR models (having the chosen kernel functions) with the batch prediction algorithm (which uses TOW as one of the predictors) being selected as the final models for fuel flow rate prediction. These models are therefore, used for all further analyses.

## 5.4 Model results and comparative analysis

In this section, the predictive performance of the chosen GPR models with batch prediction algorithm is evaluated on unseen test data. The models are compared to other APMs widely used currently for fuel flow rate estimation.

### 5.4.1 Alternative models used for comparison

The statistical models developed are compared to current state-of-the-practice APMs that are used to estimate fuel burn in tools such as the FAA's Aviation Environmental Design Tool (AEDT). These include:

- **Base of Aircraft Data (BADA) model [22]**: Developed by EUROCONTROL, BADA is a total energy-based method used for aircraft performance modeling. It uses simplified equations to model aircraft performance, with different equation coefficients (maintained in the BADA database) for different aircraft types. This method needs engine net thrust values to estimate the fuel flow rate. In this chapter, net thrust ($F_n$) values are determined using the BADA Family 3[1] thrust estimation equations for a jet engine, given by

$$F_n = \begin{cases} F_{n_{\text{as}}}, & \text{in ascent and climb out,} \\ \min(D_{\text{cr}}, F_{n_{\text{as}}}), & \text{in cruise, and} \\ F_{n_{\text{de}}}, & \text{in descent, and approach.} \end{cases} \tag{5.17}$$

The ascent thrust $F_{n_{\text{as}}}$ is given by the following set of equations:

$$F_{n_{\text{as}}} = (1 - C_{Tc,5}\Delta T_{\text{eff}})C_{Tc,1}\left(1 - \frac{h}{C_{Tc,2}} + C_{Tc,3}h^2\right) \tag{5.18}$$

$$\Delta T_{\text{eff}} = \Delta T - C_{Tc,4} \tag{5.19}$$

$$0 \leq \Delta T_{\text{eff}} \times C_{Tc,5} \leq 0.4 \tag{5.20}$$

$$C_{Tc,5} \geq 0 \tag{5.21}$$

Here, $C_{Tc,1}$, $C_{Tc,2}$, $C_{Tc,3}$, $C_{Tc,4}$, and $C_{Tc,5}$ are thrust coefficients enumerated in the BADA database, and $\Delta T$ is the temperature deviation from the ISA. $h$ is the altitude.

$D_{\text{cr}}$ in the expression for the cruise thrust in Equation 5.17 refers to the aerodynamic drag in cruise calculated using the drag coefficients for cruise ($C_{D_{0\,\text{cr}}}$ and $C_{D_{2\,\text{cr}}}$) in the BADA

[1]Revision 3.13

database.

$$D_{\mathrm{cr}} = q\mathcal{S}C_{D_{0\,\mathrm{cr}}} + \frac{C_{D_{2\,\mathrm{cr}}} m^2 g^2}{q\mathcal{S}} \tag{5.22}$$

The descent thrust $F_{n_{\mathrm{de}}}$ is given as

$$F_{n_{\mathrm{de}}} = \begin{cases} C_{\mathrm{Tdes,high}} \times F_{n_{\mathrm{as}}}, & \text{if } h > h_{\mathrm{de}}, \\ C_{\mathrm{Tdes,low}} \times F_{n_{\mathrm{as}}}, & \text{if } h < h_{\mathrm{de}} \text{ \& aircraft is in cruise config.}, \\ C_{\mathrm{Tdes,app}} \times F_{n_{\mathrm{as}}}, & \text{if } h < h_{\mathrm{de}} \text{ \& aircraft is in approach config.}, \\ C_{\mathrm{Tdes,ld}} \times F_{n_{\mathrm{as}}}, & \text{if } h < h_{\mathrm{de}} \text{ \& aircraft is in landing config.} \end{cases} \tag{5.23}$$

Here, $C_{\mathrm{Tdes,high}}$, $C_{\mathrm{Tdes,low}}$, $C_{\mathrm{Tdes,app}}$, $C_{\mathrm{Tdes,ld}}$ are descent thrust coefficients enumerated in the BADA database. $h_{\mathrm{de}}$ is the transition altitude for calculation of descent thrust and is enumerated in the BADA database. The definitions of approach and landing configurations are given in the BADA manual. These configurations are treated differently due to different aerodynamic drag values arising from the extension of slats, flaps, landing gear. In all the analysis done here, the values of the altitude, ambient temperature, ambient density, true airspeed, gross mass have been obtained from FDR data.

Once the net thrust is determined, the following equations are used to estimate the fuel flow rate per engine ($\dot{m}_{f_{\mathrm{BADA}}}$):

$$\begin{aligned} \dot{m}_{f_{\mathrm{BADA}}} &= \begin{cases} \frac{\mathrm{TSFC_{BADA}} \times F_n}{N_{\mathrm{eng}}}, & \text{in ascent, climb out, cruise,} \\ \frac{1}{N_{\mathrm{eng}}} \max(\mathrm{TSFC_{BADA}} \times F_n,\ \dot{m}_{f_{\mathrm{min}}}), & \text{in descent and approach.} \end{cases} \\ \mathrm{TSFC_{BADA}} &= C_{f1}\left(1 + \frac{V}{C_{f2}}\right) \\ \dot{m}_{f_{\mathrm{min}}} &= C_{f3}\left(1 - \frac{h}{C_{f4}}\right) \end{aligned} \tag{5.24}$$

Here, $\mathrm{TSFC_{BADA}}$ is the Thrust Specific Fuel Consumption given by BADA, $F_n$ is the aircraft net thrust from all engines, $N_{\mathrm{eng}}$ is the number of engines, $V$ is the aircraft true

airspeed, $\dot{m}_{f_{\min}}$ is the minimum aircraft fuel flow rate from all engines, and $h$ is the aircraft altitude Above Mean Sea Level (AMSL). $C_{f1}$, $C_{f2}$, $C_{f3}$, and $C_{f4}$ are aircraft type-specific constants found in the BADA database. The values of $V$ and $h$ are obtained from the FDR dataset for the analysis done here.

- **Senzig-Fleming-Iovinelli (SFI) model [28]**: The SFI model is used in AEDT to model the fuel flow rate in the terminal region of ascent and descent (< 10,000' AFE). The SFI model has been used here for performance comparison with the GPR models developed in climb out and approach (as both these phases occur below 10,000' AFE). The fuel flow rate per engine ($\dot{m}_{f_{\text{SFI}}}$) is expressed as

$$\dot{m}_{f_{\text{SFI}}} = \begin{cases} \left(K_1 + K_2 M_\infty + K_3 h + K_4 \frac{F_{n,\text{pe}}}{\delta_\infty}\right)\sqrt{\theta_\infty} F_{n,\text{pe}}, & \text{in climb out, and} \\ \left(\alpha + \beta_1 M_\infty + \beta_2 \exp\left(-\frac{\beta_3 \frac{F_{n,\text{pe}}}{\delta_\infty}}{F_{00}}\right)\right)\sqrt{\theta_\infty} F_{n,\text{pe}}, & \text{in approach.} \end{cases}$$

Here, $M_\infty$ is the aircraft Mach number, $h$ is the aircraft altitude AMSL, $F_{n,\text{pe}}$ is the aircraft net thrust per engine, $\delta_\infty$ is the ambient pressure divided by the sea level ISA reference value, $\theta_\infty$ is the ambient temperature divided by the sea level ISA reference value,- and $F_{00}$ is the maximum ISA sea level static engine thrust. $K_1, K_2, K_3, K_4, \alpha, \beta_1, \beta_2$, and $\beta_3$ are aircraft type specific model coefficients enumerated in the AEDT database. The per-engine net thrust ($F_{n,\text{pe}}$) values required as inputs to the SFI model are determined by using the BADA thrust estimation equations (that is, Equations 5.17-5.23) to calculate the net thrust from all the engines and then dividing it by the number of engines. The values of $M_\infty$, $h$, etc. are obtained from the FDR dataset.

- **ICAO Databank with Boeing Fuel Flow Method 2 Correction (ICAO-BFFM2) model [10]**: The ICAO Aircraft Engine Emissions Databank tabulates the values of fuel flow rates in the Landing and Takeoff (LTO) cycle [48]. These values are obtained through ground-based uninstalled engine certification tests and the measurements are reported after being corrected to standard sea level static ISA reference conditions. The databank entry for each engine reports four corrected values for the fuel flow rate, namely, at take-

off (100% thrust), climb out (85% thrust), approach (30% thrust), and ground idle/taxi (7% thrust). The Boeing Fuel Flow Method 2 (BFFM2) provides corrections to convert the ICAO Databank values to at-altitude conditions for an installed engine [10]. In this chapter, the ICAO Databank with BFFM2 correction has been used to estimate fuel flow rates in climb out, and approach. The fuel flow rate is then given by

$$\dot{m}_{f_{\text{ICAO-BFFM2}}} = \begin{cases} 1.013\dot{m}_{f_{\text{ICAO,co}}}\delta_\infty\theta_\infty^{-3.8}\exp\left(-0.2M_\infty^2\right), & \text{in climb out, and} \\ 1.020\dot{m}_{f_{\text{ICAO,ap}}}\delta_\infty\theta_\infty^{-3.8}\exp\left(-0.2M_\infty^2\right), & \text{in approach.} \end{cases} \tag{5.25}$$

Here, $\dot{m}_{f_{\text{ICAO-BFFM2}}}$ is the ICAO-BFFM2 fuel flow rate per engine, $\dot{m}_{f_{\text{ICAO,co}}}$ and $\dot{m}_{f_{\text{ICAO,ap}}}$ are the reported values of the fuel flow rate at reference conditions in the ICAO Databank in climb out and approach, respectively, $\delta_\infty$ and $\theta_\infty$ are the ambient pressure and temperature divided by the sea level ISA reference values, and $M_\infty$ is the aircraft Mach number. The values of the Mach number, ambient temperature, and pressure are obtained from the FDR dataset. The ICAO-BFFM2 method is used in AEDT to model the fuel flow rate in the LTO cycle when engine thrust values or other sources of the fuel flow rate are unavailable.

It is important to note that all the above APMs only provide point estimates of the fuel flow rate, and do not report any uncertainty estimates. The values of the trajectory variables, aircraft gross mass, and ambient conditions needed as inputs to calculate the fuel flow rates by the above APMs are obtained from the FDR dataset, so that the APM fuel flow rate predictive performance does not account for errors in estimating the inputs. Thus, the different models can be compared solely on the accuracy of the fuel flow rate prediction methodology.

### 5.4.2 Discussion

Tables 5.4 (a) and (b) show the ME, MAE, PC, NLPI for the GPR models of fuel flow rate (with known TOW) on unseen test data (not used for either training or validation) for different flight phases. The tables also report the ME and MAE for the BADA models in all the airborne phases, and for the SFI and ICAO-BFFM2 models in climb out and approach. Since the BADA, SFI,

ICAO-BFFM2 models report only point estimates of fuel flow rate, the PC and NLPI values for these models are not reported in the table. For the A330-243, the SFI models cannot be applied due to the unavailability of model coefficients. All the metrics are calculated on de-standardized data. Each cell in the table reports the mean and the standard deviation (within parentheses) of the evaluation metric across all the flights in the test dataset. Bold entries in the 'ME' column indicate the predicted fuel flow rates having a mean error statistically significantly different from zero (at a 5% significance level). Bold entries in the 'MAE' column indicate the model with the statistically significantly (at a 5% significance level) lowest mean absolute error among all the models compared.

It is observed that the median ME in the fuel flow rates (with known TOW) across the different aircraft types is 0.8% in ascent, 0.4% in climb out, 2.0% in cruise, 7.5% in descent, and 6.4% in approach. The majority of the aircraft types give a statistically significantly (at 5% significance level) unbiased fuel flow rate predictions in ascent, climb out, and cruise. However, there is a statistically significant bias in the predictions in descent and approach. The median MAE given by the GPR models is 4.6% in ascent, 3.8% in climb out, 10.9% in cruise, 22.4% in descent, and 18.0% in approach. The GPR models give the lowest MAE as compared to the other APMs. The GPR models achieve a reduction in median MAE of as much as 48% in ascent, 71% in climb out, 49% in cruise, 31% in descent, and 77% in approach. The median PC across the different aircraft types is 95% in ascent, 91.8% in climb out, 96.1% in cruise, 93.2% in descent, and 94.4% in approach. The closeness of these empirically observed PCs to 95% indicates good model performance. The median NLPI is 27.4% in ascent, 17.9% in climb out, 68.8% in cruise, 135.3% in descent, and 106.0% in approach. The MAE and the NLPI are the largest in descent as descent shows large operational variability inherently.

Figures 5-2–5-4 show the GPR mean predictions and 95% predictions intervals as functions of time, for one flight in the test dataset. The plots also show the predictions from the other APMs and the ground truth values of the fuel flow rates. It can be clearly observed that the mean predictions of the GPR models are more accurate than those of the other APMs. Moreover, the GPR predictions capture the variation and the uncertainty in fuel flow rates better than the other APMs. The GPR predictions are the most accurate in ascent and the least accurate in descent.

Table 5.4: (a) and (b). Predictive performance of the GPR, BADA, SFI, ICAO-BFFM2 fuel flow rate models on unseen test data. Each cell reports the mean and standard deviation (in parentheses) of the evaluation metric across all flights. All values are in percentages.

(a)

| A/C Type | Phase | ME | | | | MAE | | | | PC | NLPI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GPR | BADA | SFI | ICAO-BFFM2 | GPR | BADA | SFI | ICAO-BFFM2 | GPR | GPR |
| A319-112 | Ascent | 0.9 (2.4) | -0.8 (2.6) | - | - | **4.9 (1.9)** | 5.6 (1.9) | - | - | 96.5 (3.2) | 27.3 (1.9) |
| | Climb out | 0.1 (1.2) | **-4.2 (3.1)** | **6.5 (3.5)** | **-13.0 (5.3)** | **2.6 (1.0)** | 4.7 (3.3) | 8.2 (1.4) | 13.0 (5.2) | 95.6 (6.8) | 13.9 (1.0) |
| | Cruise | -0.5 (10.7) | **-3.1 (15.1)** | - | - | **12.1 (5.9)** | 24.7 (9.1) | - | - | 95.1 (9.9) | 68.8 (12.8) |
| | Descent | **6.7 (5.8)** | **-21.6 (8.8)** | - | - | **21.1 (4.6)** | 29.6 (6.9) | - | - | 95.6 (3.9) | 128.5 (17.4) |
| | Approach | **6.7 (8.0)** | **-12.0 (12.0)** | **38.0 (19.0)** | **65.5 (23.1)** | **17.4 (6.8)** | 24.4 (8.2) | 51.0 (9.9) | 68.6 (21.9) | 94.7 (4.9) | 95.8 (13.5) |
| A320-214 | Ascent | 0.7 (2.0) | -0.8 (2.7) | - | - | **3.8 (1.6)** | 5.8 (2.0) | - | - | 95.3 (5.0) | 20.7 (2.4) |
| | Climb out | 0.4 (2.1) | **-4.7 (3.7)** | **6.4 (4.1)** | **-2.2 (6.0)** | **3.5 (2.5)** | 7.3 (4.2) | 9.7 (2.8) | 8.6 (4.1) | 94.7 (9.8) | 20.7 (1.2) |
| | Cruise | 1.8 (15.1) | **-6.4 (8.6)** | - | - | **14.1 (9.3)** | 21.5 (5.0) | - | - | 94.7 (13.7) | 69.8 (12.4) |
| | Descent | **9.5 (8.1)** | **-14.7 (11.1)** | - | - | **22.5 (5.9)** | 32.3 (5.3) | - | - | 92.9 (3.2) | 108.0 (11.4) |
| | Approach | **6.1 (7.9)** | **4.8 (12.5)** | **73.8 (18.0)** | **80.7 (21.5)** | **16.2 (5.1)** | 24.9 (5.9) | 75.5 (16.3) | 82.2 (20.5) | 94.8 (4.0) | 92.8 (14.0) |
| A321-111 | Ascent | 0.8 (2.3) | **7.1 (3.4)** | - | - | **3.6 (1.8)** | 10.3 (2.6) | - | - | 94.3 (5.5) | 21.7 (2.5) |
| | Climb out | -0.1 (2.5) | 0.5 (4.8) | **8.0 (5.2)** | 2.4 (7.1) | **4.3 (1.7)** | 11.8 (3.5) | 14.6 (2.6) | 13.1 (4.1) | 91.7 (9.0) | 21.7 (1.2) |
| | Cruise | 2.8 (7.3) | **6.3 (12.0)** | - | - | **10.5 (5.0)** | 23.2 (8.4) | - | - | 98.8 (4.3) | 78.1 (13.6) |
| | Descent | **8.3 (6.5)** | **-21.1 (11.1)** | - | - | **22.2 (4.7)** | 33.0 (7.3) | - | - | 90.8 (4.6) | 109.4 (18.5) |
| | Approach | **5.5 (6.3)** | **-9.5 (15.2)** | **35.3 (19.2)** | **56.7 (27.9)** | **16.4 (4.4)** | 25.7 (7.4) | 43.3 (14.8) | 58.9 (26.0) | 92.5 (5.4) | 77.6 (16.0) |
| A330-202 | Ascent | -0.0 (1.8) | **-7.7 (2.3)** | - | - | **4.5 (2.2)** | 9.3 (1.4) | - | - | 96.7 (3.2) | 27.5 (2.6) |
| | Climb out | **2.6 (2.1)** | **-10.2 (1.9)** | **-4.7 (2.1)** | **-17.3 (2.9)** | **4.1 (1.8)** | 11.1 (1.6) | 10.2 (1.4) | 17.3 (2.9) | 91.8 (9.7) | 18.3 (1.3) |
| | Cruise | -1.4 (8.1) | -1.8 (4.1) | - | - | **9.4 (3.8)** | 13.6 (2.9) | - | - | 94.4 (16.3) | 60.5 (19.3) |
| | Descent | **9.4 (6.8)** | **21.8 (14.7)** | - | - | **27.7 (3.8)** | 45.2 (12.2) | - | - | 92.1 (4.9) | 176.5 (21.5) |
| | Approach | **10.0 (14.1)** | **27.4 (13.0)** | **108.3 (23.3)** | **88.3 (30.2)** | **30.0 (11.2)** | 46.9 (13.8) | 113.0 (22.6) | 98.0 (25.8) | 91.2 (7.5) | 146.3 (26.5) |

(b)

| A/C Type | Phase | ME | | | | MAE | | | | PC | NLPI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GPR | BADA | SFI | ICAO-BFFM2 | GPR | BADA | SFI | ICAO-BFFM2 | GPR | GPR |
| A330-243 | Ascent | 0.4 (2.2) | **-6.2 (3.5)** | - | - | **4.0 (1.7)** | 8.3 (2.0) | - | - | 94.7 (10.3) | 26.1 (6.4) |
| | Climb out | 0.8 (2.0) | **-9.9 (2.0)** | - | -0.5 (6.0) | **3.1 (2.2)** | 10.3 (1.7) | - | 5.9 (3.2) | 91.2 (18.4) | 12.9 (0.9) |
| | Cruise | **6.0 (9.3)** | 1.9 (7.3) | - | - | **11.0 (8.2)** | 18.4 (9.1) | - | - | 96.5 (4.2) | 53.6 (17.5) |
| | Descent | **11.7 (11.7)** | -2.7 (13.5) | - | - | **23.3 (10.8)** | 32.6 (5.7) | - | - | 93.8 (6.3) | 143.6 (22.7) |
| | Approach | 9.9 (12.2) | 12.3 (20.1) | - | 95.7 (26.7) | **20.4 (10.3)** | 36.5 (6.5) | - | 96.8 (25.8) | 91.5 (8.1) | 111.4 (23.1) |
| A340-541 | Ascent | 1.3 (3.1) | 0.9 (3.7) | - | - | **4.7 (1.8)** | 9.5 (2.5) | - | - | 93.4 (6.8) | 31.9 (12.6) |
| | Climb out | 1.8 (5.0) | **10.5 (6.8)** | **10.5 (5.9)** | 8.6 (16.6) | **4.7 (3.3)** | 11.7 (6.1) | 11.0 (5.6) | 13.1 (14.0) | 78.5 (28.5) | 14.6 (0.9) |
| | Cruise | 0.3 (13.1) | **8.7 (4.7)** | - | - | **11.3 (6.8)** | 21.0 (5.7) | - | - | 96.1 (8.9) | 68.8 (39.4) |
| | Descent | 1.3 (11.1) | **-19.8 (13.1)** | - | - | **22.3 (6.1)** | 30.6 (9.5) | - | - | 93.4 (3.4) | 142.0 (26.7) |
| | Approach | 1.8 (5.8) | **-18.0 (16.2)** | **27.6 (23.4)** | **56.3 (25.0)** | **16.6 (2.6)** | 29.0 (10.2) | 42.3 (14.0) | 63.1 (20.1) | 95.9 (3.5) | 126.0 (24.2) |
| B767-300 | Ascent | **0.8 (2.6)** | **-3.1 (2.5)** | - | - | **4.6 (1.9)** | 6.6 (2.2) | - | - | 97.0 (4.3) | 29.0 (2.8) |
| | Climb out | 0.4 (1.5) | **-5.8 (1.1)** | **-8.9 (1.4)** | **-21.0 (4.4)** | **2.5 (0.9)** | 5.8 (1.1) | 8.9 (1.4) | 21.0 (4.4) | 98.4 (3.5) | 17.4 (1.1) |
| | Cruise | 3.3 (10.2) | **6.1 (13.8)** | - | - | **10.8 (8.7)** | 21.4 (13.9) | - | - | 96.9 (7.1) | 77.2 (28.0) |
| | Descent | **6.4 (7.6)** | 0.9 (18.4) | - | - | **23.1 (5.3)** | 34.6 (6.2) | - | - | 93.7 (5.7) | 144.7 (43.9) |
| | Approach | **5.8 (9.7)** | 6.3 (20.3) | **35.7 (23.8)** | **65.9 (33.2)** | **19.3 (7.1)** | 33.6 (6.3) | 49.0 (13.2) | 78.7 (29.3) | 94.9 (6.1) | 126.7 (37.7) |
| B777-300 ER | Ascent | 0.1 (2.6) | -4.6 (6.5) | - | - | **6.7 (2.5)** | 13.8 (4.2) | - | - | 93.3 (5.2) | 35.8 (5.1) |
| | Climb out | -0.0 (4.2) | **-19.1 (8.8)** | **-9.7 (10.0)** | **-20.0 (8.4)** | **6.7 (1.7)** | 23.0 (5.5) | 21.0 (5.8) | 23.1 (5.7) | 90.5 (11.6) | 32.2 (6.7) |
| | Cruise | **2.1 (4.4)** | **6.2 (3.9)** | - | - | **9.3 (5.4)** | 19.9 (7.5) | - | - | 96.1 (5.3) | 49.1 (12.3) |
| | Descent | **3.4 (5.8)** | **-4.5 (8.5)** | - | - | **17.2 (3.7)** | 25.1 (6.6) | - | - | 91.8 (5.0) | 122.1 (13.8) |
| | Approach | **7.4 (8.8)** | **-7.8 (12.0)** | **28.5 (14.4)** | **66.4 (26.4)** | **18.5 (4.7)** | 23.5 (7.3) | 39.1 (9.2) | 76.2 (23.6) | 94.0 (4.7) | 100.6 (16.8) |

An examination of the ground truth fuel flow rates reveals that unlike ascent which is a smooth procedure, descent undergoes larger procedural variability. This variability is reflected in poorer mean predictions and larger prediction intervals. Despite the poorer predictions in descent, the GPR models are able to capture the time scales in fuel flow rates during descent better than the other APMs.

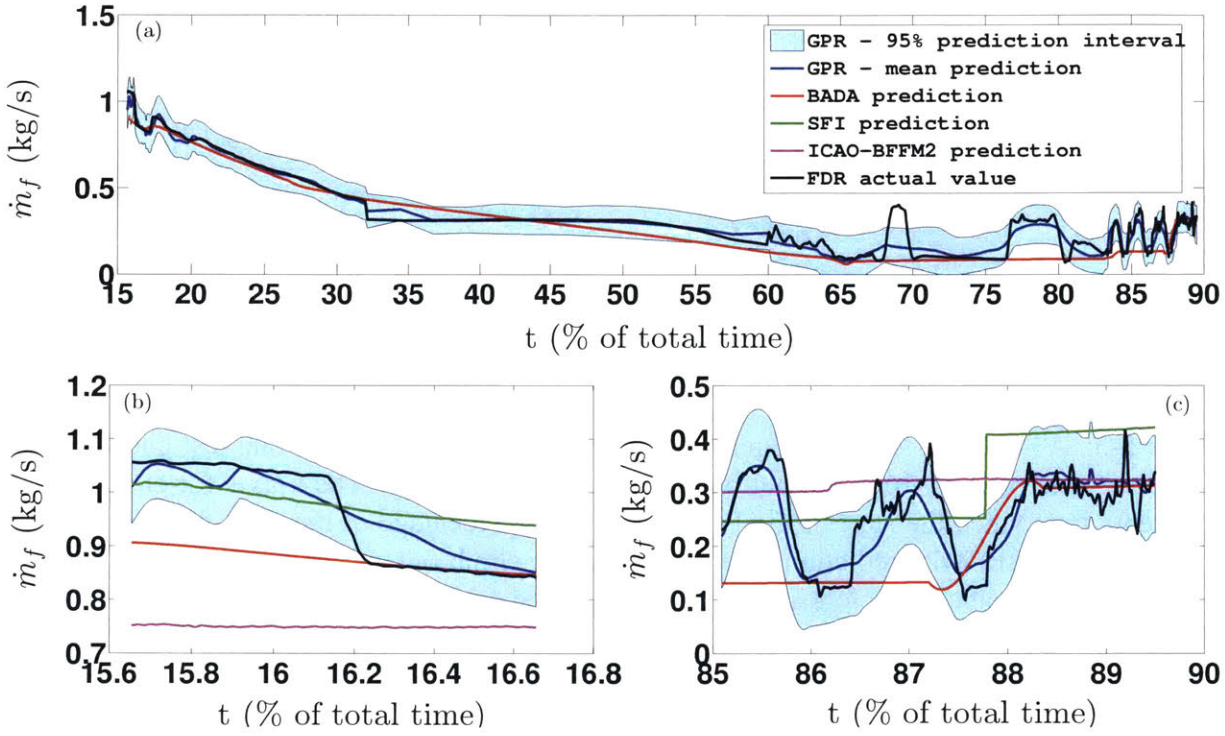



Figure 5-2: A319-112: Averaged fuel flow rate profile (per engine) for one test data flight in (a) all airborne phases, (b) climb out only, and (c) approach only. The x-axis for each subplot represents time since the start of the FDR record as a percentage of the total flight time. The SFI and ICAO-BFFM2 predictions are present in the climb out and approach phases only. It should be noted that all subplots are on de-standardized data and have different scales.

## 5.5 Cumulative fuel burn by airborne phase

The predicted fuel flow rate profiles can be further used to predict the cumulative fuel burn (total mass of fuel consumed by all the engines) in each airborne phase. The equation for predicting

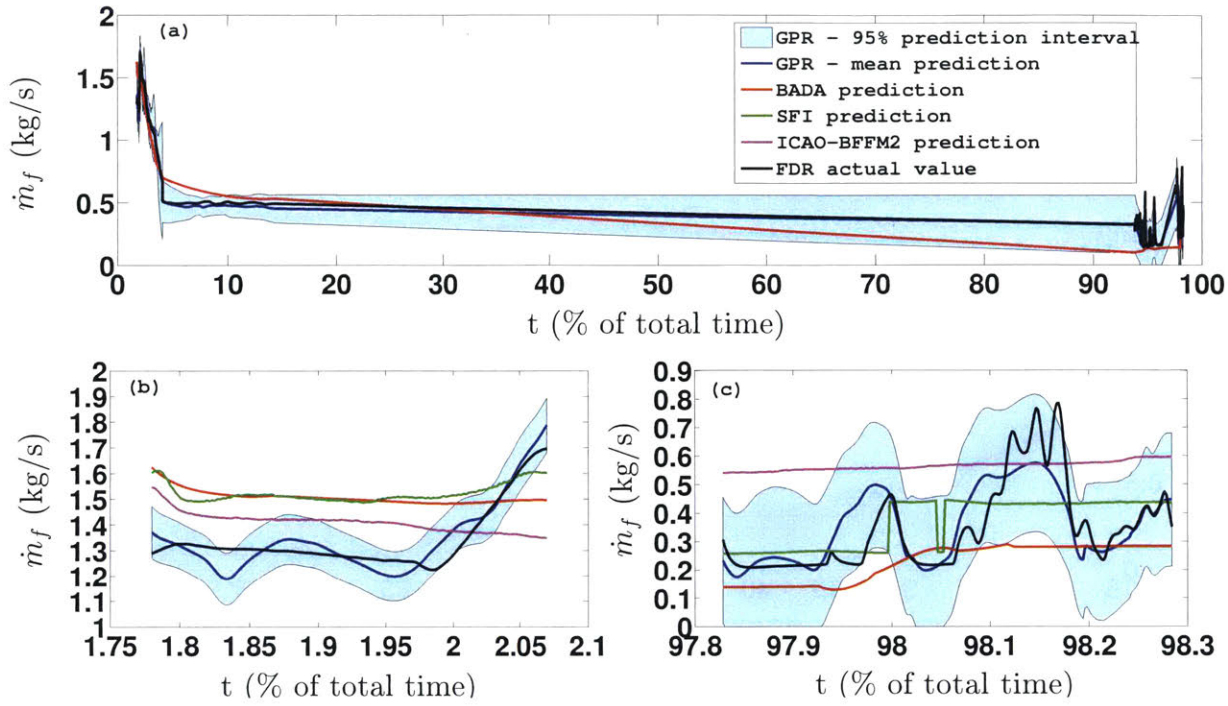



Figure 5-3: A340-541: Averaged fuel flow rate profile (per engine) for one test data flight in (a) all airborne phases, (b) climb out only, and (c) approach only. The x-axis for each subplot represents time since the start of the FDR record as a percentage of the total flight time. The SFI and ICAO-BFFM2 predictions are present in the climb out and approach phases only. It should be noted that all subplots are on de-standardized data and have different scales.

the cumulative fuel burn in a particular phase of flight is given by

$$\hat{m}_f = N_{\text{eng}} \sum_{i=1}^{n-1} \hat{\dot{m}}_{f_i} \Delta t_{i+1}. \tag{5.26}$$

Here, $\hat{m}_f$ is the predicted cumulative fuel burn in a particular phase of flight, $N_{\text{eng}}$ is the number of engines on the aircraft, $n$ is the number of time instants in that phase of flight, $\hat{\dot{m}}_{f_i}$ is the predicted fuel flow rate per engine for the $i^{\text{th}}$ instant, and $\Delta t_{i+1} = t_{i+1} - t_i$ is the time interval between the $(i+1)^{\text{th}}$ and the $i^{\text{th}}$ instants. At every instant, the fuel flow rate per engine predicted by a GPR model follows a Gaussian distribution ($\mathcal{N}$). Hence, the cumulative mass of fuel consumed divided by the number of engines and the total time in phase follows a Gaussian Mixture distribution ($\mathcal{GM}$) with mixture component weights governed by the time intervals

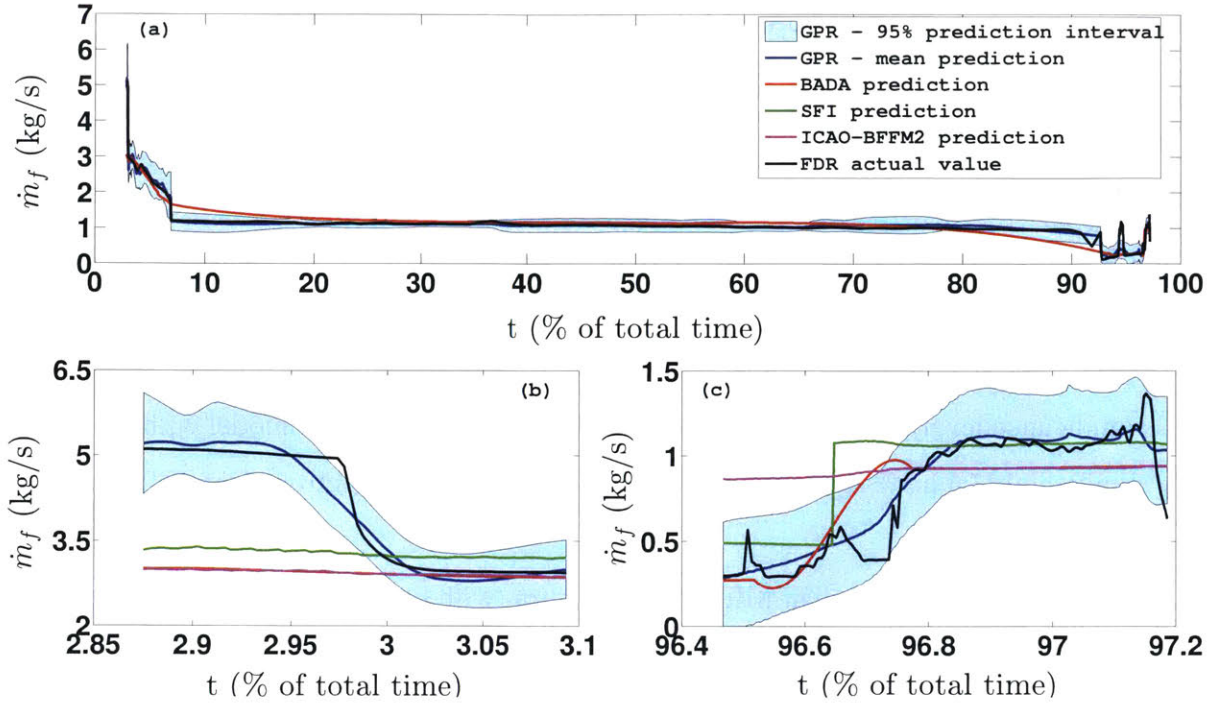



Figure 5-4: B777-300ER: Averaged fuel flow rate profile (per engine) for one test data flight in (a) all airborne phases, (b) climb out only, and (c) approach only. The x-axis for each subplot represents time since the start of the FDR record as a percentage of the total flight time. The SFI and ICAO-BFFM2 predictions are present in the climb out and approach phases only. It should be noted that all subplots are on de-standardized data and have different scales.

between successive instants.

$$\hat{\dot{m}}_{f_i} \sim \mathcal{N}(\mu_i, \sigma_i^2) \tag{5.27}$$

$$\frac{\hat{m}_f}{N_{\text{eng}}\Delta t_{\text{tot}}} \sim \mathcal{GM}\left(\boldsymbol{\mu}, \boldsymbol{\sigma^2}, \frac{\boldsymbol{\Delta t}}{\Delta t_{\text{tot}}}\right) \tag{5.28}$$

Here, $\boldsymbol{\mu}$ is the vector of the means of the GPR-predicted fuel flow rate predictive distributions for the first $n-1$ instants (the $\mu_i$s), $\boldsymbol{\sigma^2}$ is the vector of the variances of the GPR-predicted fuel flow rate predictive distributions for the first $n-1$ instants (the $\sigma_i^2$s), $\boldsymbol{\Delta t}$ is the vector of the time intervals for the last $n-1$ instants (the $\Delta t_i$s), and $\Delta t_{\text{tot}} = \sum_{i=1}^{n-1} \Delta t_{i+1}$ is the total time in phase over which the cumulative fuel burn is desired. For each flight and phase of flight, samples are drawn at random from the Gaussian Mixture distribution. The mean of these samples is used to predict the average fuel burn for the particular flight and phase of flight. The 95% prediction

interval for the predicted fuel burn is determined by using the 95% Highest Density Interval (HDI) of the Gaussian Mixture Distribution.

Tables 5.5 (a) and (b) show the predictive performance of the cumulative fuel burn models (with known TOW) on the entire unseen dataset (i.e., the combined validation and test sets) for the different aircraft types in the different airborne phases. Predictions of fuel mass consumption using other APMs are also tabulated. Bold entries in the 'ME' column indicate the predicted fuel burn values having a mean error statistically significantly different from zero (at a 5% significance level). Bold entries in the 'MAE' column indicate the model with the statistically significantly (at a 5% significance level) lowest mean absolute error among all the models compared.

It is observed that the median ME in the fuel burn (with known TOW) across the different aircraft types is 0.4% in ascent, 0.3% in climb out, -0.1% in cruise, 3.7% in descent, and 2.0% in approach. The median MAE given by the GPR models is 2.2% in ascent, 2.0% in climb out, 7.2% in cruise, 8.2% in descent, and 5.5% in approach. In the majority of the comparisons, the GPR models give the lowest MAE as compared to the other APMs. The GPR models achieve a reduction in median MAE of as much as 54% in ascent, 80% in climb out, 34% in cruise, 66% in descent, and 89% in approach. The median PC across the different aircraft types is seen to be 100% in all the phases. The median NLPI is 81.3% in ascent, 30.6% in climb out, 79.8% in cruise, 225.9% in descent, and 152.0% in approach. A 100% PC and the large NLPI indicate overly cautious prediction intervals and that smaller intervals may be sufficient to capture the uncertainty in fuel burn. However, these large intervals are not surprising and are a result of the model setup. In this thesis, the primary variable which is modeled directly is the fuel flow rate and not the mass of fuel consumed. The mass of fuel consumed (i.e., the fuel burn) is modeled indirectly via the fuel flow rate predictions. GPR models for fuel flow rate prediction have been shown to give a PC close to 95% and are hence, properly specified. However, when the estimates of fuel flow rates are combined to give the fuel burn, the individual uncertainties in the fuel flow rates at each instant also combine to give larger-than-necessary prediction intervals for the fuel burn. An alternative to predict the fuel burn more precisely would be to directly determine the mapping between the trajectory inputs and the cumulative fuel burn in different

Table 5.5: (a) and (b). Predictive performance of the GPR, BADA, SFI, ICAO-BFFM2 cumulative fuel burn models on unseen (combined validation and test) data. Each cell reports the mean and standard deviation (in parentheses) of the evaluation metric across flights. All values are in percentages.

(a)

| A/C Type | Phase | ME | | | | MAE | | | | PC | NLPI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GPR | BADA | SFI | ICAO-BFFM2 | GPR | BADA | SFI | ICAO-BFFM2 | GPR | GPR |
| A319-112 | Ascent | **1.1** (**3.3**) | -0.8 (3.6) | - | - | 2.5 (2.4) | 3.0 (2.0) | - | - | 100.0 (0.0) | 80.3 (16.1) |
| | Climb out | -0.1 (1.8) | **-4.3** (**3.5**) | **6.1** (**3.9**) | **-12.7** (**4.9**) | **1.3** (**1.3**) | 4.3 (3.5) | 6.3 (3.6) | 12.7 (4.9) | 100.0 (0.0) | 21.8 (3.6) |
| | Cruise | -1.0 (10.4) | 6.0 (18.2) | - | - | 7.8 (7.0) | 11.0 (15.6) | - | - | 100.0 (0.0) | 82.4 (17.4) |
| | Descent | -1.6 (9.0) | **-38.4** (**12.5**) | - | - | **7.6** (**5.0**) | 38.5 (12.3) | - | - | 100.0 (0.0) | 190.1 (23.3) |
| | Approach | 0.7 (5.3) | **-12.2** (**12.5**) | **42.8** (**31.1**) | **53.3** (**36.4**) | **3.9** (**3.6**) | 13.7 (10.8) | 42.8 (31.1) | 53.3 (36.4) | 100.0 (0.0) | 133.6 (16.5) |
| A320-214 | Ascent | **1.2** (**3.0**) | -0.2 (4.1) | - | - | 2.6 (1.9) | 2.9 (2.9) | - | - | 100.0 (0.0) | 82.3 (15.5) |
| | Climb out | 0.0 (1.8) | **-5.3** (**4.4**) | **5.6** (**4.8**) | **-2.8** (**6.7**) | **1.4** (**1.1**) | 5.3 (4.4) | 6.8 (2.8) | 5.9 (4.2) | 100.0 (0.0) | 33.7 (5.6) |
| | Cruise | -0.7 (12.0) | 0.0 (8.4) | - | - | 9.0 (7.9) | **6.4** (**5.3**) | - | - | 100.0 (0.0) | 77.3 (14.3) |
| | Descent | **4.5** (**9.9**) | **-39.8** (**14.3**) | - | - | **8.1** (**7.2**) | 39.8 (14.3) | - | - | 100.0 (0.0) | 177.9 (19.5) |
| | Approach | 1.7 (6.8) | 3.3 (16.6) | **63.5** (**21.7**) | **55.9** (**20.1**) | **5.5** (**4.4**) | 13.8 (9.7) | 63.5 (21.7) | 55.9 (20.1) | 100.0 (0.0) | 133.6 (18.4) |
| A321-111 | Ascent | 0.4 (3.0) | **7.5** (**4.4**) | - | - | **2.2** (**2.0**) | 7.5 (4.4) | - | - | 100.0 (0.0) | 72.8 (14.6) |
| | Climb out | -0.2 (2.6) | -0.6 (6.2) | **6.7** (**6.5**) | 1.5 (9.0) | **2.1** (**1.5**) | 5.2 (3.2) | 7.5 (5.6) | 7.2 (5.5) | 97.6 (15.6) | 41.1 (8.1) |
| | Cruise | 0.5 (9.7) | **17.9** (**20.9**) | - | - | **7.1** (**6.6**) | 19.0 (19.9) | - | - | 100.0 (0.0) | 93.3 (25.4) |
| | Descent | 1.5 (8.3) | **-40.6** (**14.4**) | - | - | **6.9** (**4.6**) | 40.6 (14.4) | - | - | 100.0 (0.0) | 171.8 (25.9) |
| | Approach | **2.4** (**5.6**) | **-12.3** (**15.0**) | **30.6** (**17.8**) | **45.8** (**23.0**) | **4.9** (**3.4**) | 15.5 (11.5) | 31.3 (16.6) | 45.8 (23.0) | 100.0 (0.0) | 114.3 (21.6) |
| A330-202 | Ascent | **-0.8** (**1.9**) | **-7.3** (**2.0**) | - | - | **1.6** (**1.3**) | 7.3 (2.0) | - | - | 100.0 (0.0) | 71.4 (13.7) |
| | Climb out | **2.2** (**1.7**) | **-11.2** (**2.0**) | **-5.8** (**2.1**) | **-18.1** (**3.1**) | **2.2** (**1.7**) | 11.2 (2.0) | 5.8 (2.1) | 18.1 (3.1) | 93.1 (25.8) | 36.8 (4.4) |
| | Cruise | -0.8 (7.3) | -0.4 (3.7) | - | - | 5.8 (4.3) | **3.0** (**2.2**) | - | - | 100.0 (0.0) | 70.5 (14.3) |
| | Descent | **9.5** (**11.3**) | **-10.0** (**18.0**) | - | - | 11.6 (9.1) | 16.6 (12.0) | - | - | 100.0 (0.0) | 236.1 (19.7) |
| | Approach | 2.6 (8.1) | **8.1** (**14.1**) | **82.9** (**42.7**) | **47.5** (**42.9**) | **6.2** (**5.7**) | **11.8** (**11.2**) | 82.9 (42.7) | 47.8 (42.6) | 100.0 (0.0) | 188.9 (22.5) |

(b)

| A/C Type | Phase | ME | | | | MAE | | | | PC | NLPI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GPR | BADA | SFI | ICAO-BFFM2 | GPR | BADA | SFI | ICAO-BFFM2 | GPR | GPR |
| A330-243 | Ascent | 0.2 (2.8) | **-4.3 (4.3)** | - | - | **1.9 (2.0)** | 5.3 (2.8) | - | - | 100.0 (0.0) | 67.0 (14.8) |
| | Climb out | **0.9 (2.0)** | **-10.1 (2.1)** | - | -0.7 (5.9) | **1.8 (1.2)** | 10.1 (2.1) | - | 4.8 (3.3) | 100.0 (0.0) | 18.0 (2.9) |
| | Cruise | **3.4 (8.1)** | **6.0 (10.5)** | - | - | 5.5 (6.8) | 6.7 (10.1) | - | - | 100.0 (0.0) | 64.7 (15.7) |
| | Descent | **8.0 (10.1)** | **-18.5 (17.8)** | - | - | **10.4 (7.5)** | 19.7 (16.6) | - | - | 100.0 (0.0) | 218.4 (23.0) |
| | Approach | **6.6 (9.8)** | 6.1 (18.0) | - | **85.8 (53.8)** | **7.4 (9.2)** | 14.7 (11.9) | - | 85.8 (53.8) | 100.0 (0.0) | 150.7 (23.5) |
| A340-541 | Ascent | -0.0 (2.1) | **-4.4 (3.2)** | - | - | **1.6 (1.3)** | 4.7 (2.7) | - | - | 100.0 (0.0) | 92.6 (17.6) |
| | Climb out | 1.0 (3.9) | **9.5 (6.3)** | **9.8 (5.9)** | **5.3 (12.9)** | **2.6 (3.1)** | 9.9 (5.7) | 10.0 (5.3) | 7.4 (11.8) | 100.0 (0.0) | 27.4 (7.1) |
| | Cruise | -4.5 (13.7) | **11.3 (6.8)** | - | - | 9.7 (10.5) | 11.3 (6.8) | - | - | 100.0 (0.0) | 82.2 (54.2) |
| | Descent | 1.9 (10.8) | **-28.1 (10.2)** | - | - | **8.3 (6.9)** | 28.1 (10.2) | - | - | 100.0 (0.0) | 233.4 (24.9) |
| | Approach | -2.4 (5.9) | **-23.8 (14.4)** | **55.2 (89.8)** | **72.4 (101.1)** | **5.4 (3.2)** | 24.5 (13.0) | 55.2 (89.8) | 72.6 (101.0) | 100.0 (0.0) | 160.4 (23.3) |
| B767-300 | Ascent | **1.1 (2.9)** | **-2.2 (2.8)** | - | - | 2.3 (1.9) | 3.1 (1.8) | - | - | 100.0 (0.0) | 84.8 (14.1) |
| | Climb out | 0.5 (1.9) | -5.2 (1.4) | -8.3 (1.8) | -21.0 (3.6) | **1.4 (1.3)** | 5.2 (1.4) | 8.3 (1.8) | 21.0 (3.6) | 100.0 (0.0) | 23.7 (4.6) |
| | Cruise | 4.5 (11.5) | **15.6 (25.2)** | - | - | **7.3 (9.9)** | 15.9 (25.1) | - | - | 100.0 (0.0) | 92.4 (30.7) |
| | Descent | **6.4 (8.9)** | **-15.3 (14.1)** | - | - | 9.2 (5.8) | 16.5 (12.6) | - | - | 100.0 (0.0) | 255.9 (20.3) |
| | Approach | 2.0 (5.8) | -0.6 (17.0) | **23.3 (24.0)** | **30.5 (30.0)** | **4.6 (4.0)** | 13.3 (10.2) | 25.7 (21.3) | 31.3 (29.0) | 100.0 (0.0) | 167.6 (28.4) |
| B777-300 ER | Ascent | **0.3 (3.1)** | **-2.5 (5.2)** | - | - | **2.2 (2.2)** | 4.9 (3.0) | - | - | 100.0 (0.0) | 83.2 (14.6) |
| | Climb out | -0.5 (3.6) | **-22.3 (6.5)** | **-12.9 (7.4)** | **-22.1 (7.1)** | **2.9 (2.1)** | 22.3 (6.5) | 13.2 (7.0) | 22.1 (7.1) | 73.9 (44.4) | 79.4 (9.0) |
| | Cruise | 1.3 (7.9) | **10.7 (16.1)** | - | - | **4.8 (6.4)** | 10.8 (16.1) | - | - | 100.0 (0.0) | 58.2 (17.5) |
| | Descent | **2.9 (8.6)** | **-6.3 (12.3)** | - | - | 7.4 (5.1) | 9.9 (9.6) | - | - | 100.0 (0.0) | 255.0 (14.7) |
| | Approach | 2.0 (7.0) | **-12.3 (11.0)** | **20.7 (34.9)** | **37.8 (37.3)** | **5.9 (4.2)** | 13.2 (9.9) | 22.1 (34.0) | 38.1 (37.0) | 100.0 (0.0) | 153.2 (22.4) |

phases. This approach has not been pursued further in this thesis, and is a potential direction for future research.

## 5.6 Total airborne fuel burn

Table 5.6 shows the ME and the MAE for the total fuel burn prediction in the air (with known TOW). The total fuel burn in the air is predicted by adding the individual fuel burn predictions in ascent, cruise, and descent. The table also shows the BADA predictive performance for the total fuel burn in the air. Statistical tests are used to compare the MEs with zero and MEs that are statistically significantly different from zero (at a 5% significance level) are highlighted in bold. Statistical tests are also used to compare the MAEs of the GPR models to those of the BADA models. The model which gives a statistically significantly lower MAE is highlighted in bold in the table.

Table 5.6: Performance of the GPR, and the BADA models to predict the total airborne fuel burn for unseen (combined validation and test) data. Each cell reports the mean and standard deviation (within parentheses) of the evaluation metric across all the flights.

| A/C Type | ME (%) | | MAE (%) | |
|---|---|---|---|---|
| | GPR | BADA | GPR | BADA |
| A319-112 | -0.7 (4.7) | **-8.9 (5.7)** | **3.9 (2.7)** | 9.0 (5.6) |
| A320-214 | 0.4 (5.1) | **-8.7 (7.3)** | **4.0 (3.2)** | 9.4 (6.4) |
| A321-111 | -0.0 (4.1) | -2.5 (11.3) | **3.4 (2.2)** | 9.2 (6.8) |
| A330-202 | -0.5 (5.3) | **-3.1 (2.9)** | 4.2 (3.2) | 3.5 (2.4) |
| A330-243 | **1.7 (3.7)** | -1.4 (4.2) | 3.2 (2.5) | 3.0 (3.3) |
| A340-541 | -1.3 (7.1) | **2.3 (3.1)** | 5.4 (4.7) | **2.7 (2.7)** |
| B767-300 | **2.0 (4.7)** | **3.0 (4.6)** | 3.7 (3.4) | 4.1 (3.6) |
| B777-300ER | 0.5 (3.3) | **5.1 (2.4)** | **2.8 (1.9)** | 5.2 (2.2) |

Figures 5-5 and 5-6 contain box-plots showing the errors and the absolute errors, respectively, for the total airborne fuel burn prediction across the different flights in unseen data not used for training. Model predictions are unbiased (at a 5% significance level) if the error box-plots in Figure 5-5 contain zero in their notched regions. The notched regions of the absolute error box-plots in Figure 5-6 for the GPR models can be compared to those of the box-plots for

the corresponding BADA models to determine which model gives a statistically significantly lower absolute error (at a 5% significance level) on the total airborne fuel burn prediction.

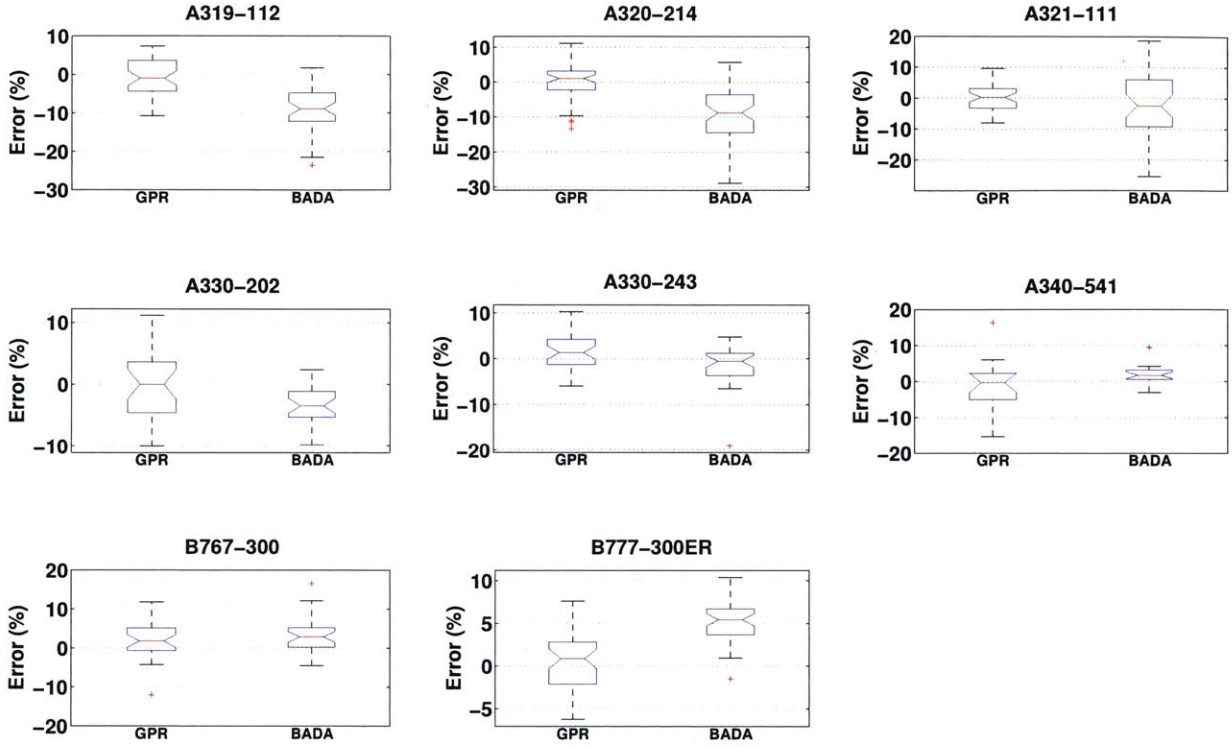



Figure 5-5: Box-plots showing error on total airborne fuel burn prediction for flights in unseen data not used for training. On each box, the central mark is the median, the edges of the box are the $25^{th}$ and the $75^{th}$ percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted as red crosses. The notch in the box represents an interval which can be used for statistical comparison at a 5% significance level. These box-plots are helpful in checking for model bias in the mean predictions.

Table 5.6 shows that the median of the total airborne fuel burn ME (with known TOW) for the GPR models across the different aircraft types is 0.2%. For the majority of the aircraft types, the GPR models give unbiased predictions of the total airborne fuel burn. The median MAE (with known TOW) for the GPR and the BADA models is 3.8% and 4.7%, respectively. For all the aircraft types except the A340-541, the GPR models give a statistically significantly similar or lower MAE as compared to the BADA models. In other words, the GPR models provide better predictions of the total airborne fuel burn as compared to the BADA models. Similar to the case of predicting the cumulative fuel burn by phase, it may be possible to determine more

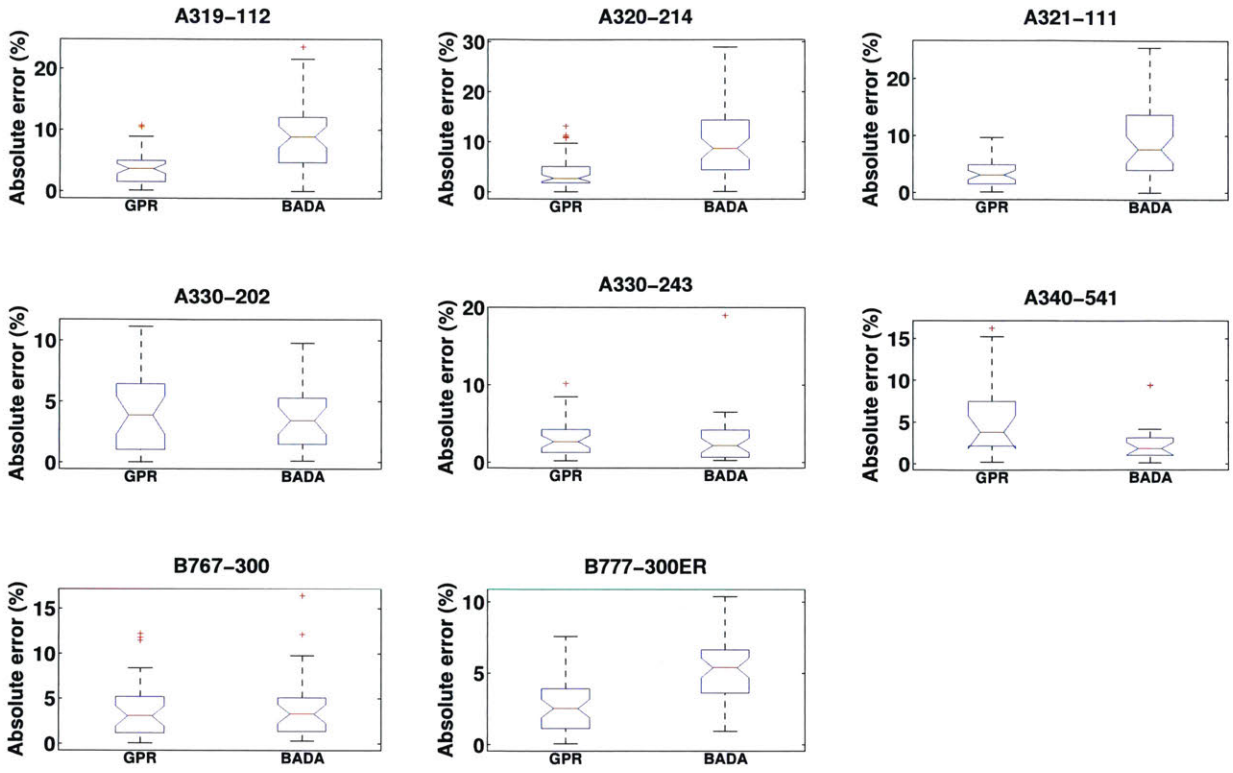



Figure 5-6: Box-plots showing absolute error on total airborne fuel burn prediction for flights in unseen data not used for training. On each box, the central mark is the median, the edges of the box are the $25^{\text{th}}$ and the $75^{\text{th}}$ percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted as red crosses. The notch in the box represents an interval which can be used for statistical comparison at a 5% significance level. These box-plots can be used to compare the GPR models to the corresponding BADA models for the accuracy of the mean predictions.

accurate predictions of the total airborne fuel burn by treating it as the predicted variable, rather than aggregating the instantaneous fuel flow rates. The development of such models of total fuel burn is an interesting topic for future research.

A comment regarding the trajectory-based input variables used for model testing is in order. In this chapter, models have been primarily evaluated for their fuel flow and fuel burn predictive performance without accounting for possible errors in the trajectory-based input variables. The actual values of the input variables as enumerated in the FDR dataset are therefore, used to estimate the fuel flow rates for flights in the unseen dataset not used for training. In reality, trajectory-based input variables for a new flight will be obtained from ground-based surveillance

systems and the values of these variables may differ from those in the FDR dataset for that flight. Future work would also look at evaluating model predictive performance in the airborne phases using inputs from surveillance systems, such as the Traffic Flow Management System (TFMS) in the United States.

## 5.7 Intra-family model generalizability

In the previous sections, each aircraft and engine variant combination has a separate model to estimate the fuel flow rates. In this section, the potential of the models developed for one aircraft/engine type to generalize to other aircraft/engine types is examined. A generalized model, if possible, is practically advantageous as the same model can be used to estimate fuel flow rates for different aircraft types, thereby reducing the total number of models required to be developed.

The first step towards checking for model generalizability is to examine how well models trained for one aircraft type predict the fuel flow rates for other aircraft types within the same family. Two aircraft families can be identified in the FDR-I dataset - the A320 family comprising the A319-112, the A320-214, and the A321-111 aircraft types, and the A330-200 family comprising the A330-202 and the A330-243 aircraft types. The A320-214 and the A330-243 are chosen as the representative aircraft types for these two families. Models trained on these two aircraft are used to estimate fuel flow rates for the flights in the test datasets of the other aircraft types in the same family. Good intra-family generalizability would imply that the resultant fuel flow rate estimates are statistically significantly similar to those obtained in Section 5.4 by using different aircraft type-specific models.

Table 5.7 shows model evaluation metrics on test data when fuel flow rates are estimated using both the aircraft type-specific and the family generalized models. The GPR models trained for the A320-214 in Section 5.3.1 serve as the generalized models for the A319-112 and the A321-111 in the A320 family, while the GPR models trained for the A330-243 in Section 5.3.1 serve as the generalized models for the A330-202 in the A330-200 family. The aircraft type-specific ('Specific') and the family generalized ('General') models are statistically compared to

one another. In these comparisons, models with a statistically significantly (at a 5% significance level) lower MAE, higher PC, and lower NLPI are highlighted in bold in the table. From Table 5.7, it can be seen that barring a few cases (such as some phases for the A319-112), models trained on an aircraft type do not seem to be generalizable, in terms of the fuel flow rate prediction, to the other aircraft types within the same family. Future work could look into ways in which models can be made more generalizable. One way to improve model generalizability could be to combine data from different aircraft types of a family to train representative models for that family.

Table 5.7: Examination of intra-family model generalizability: Model evaluation metrics when aircraft type-specific ('Specific') and family generalized ('General') GPR models are used for fuel flow rate estimation on test data. Each cell reports the mean and standard deviation (in parentheses) of the evaluation metric across flights.

| A/C Type | Phase | MAE (%) | | PC (%) | | NLPI (%) | |
|---|---|---|---|---|---|---|---|
| | | **Specific** | **General** | **Specific** | **General** | **Specific** | **General** |
| A319-112 | Ascent | 4.9 (1.9) | 5.1 (2.0) | 96.5 (3.2) | 95.8 (4.2) | **27.3 (1.9)** | 29.9 (2.5) |
| | Climb out | **2.6 (1.0)** | 3.0 (1.0) | 95.6 (6.8) | 99.1 (3.9) | **13.9 (1.0)** | 24.0 (2.6) |
| | Cruise | 12.1 (5.9) | 11.5 (8.2) | 95.1 (9.9) | 92.2 (11.3) | 68.8 (12.8) | **54.6 (7.9)** |
| | Descent | 21.1 (4.6) | 21.8 (5.2) | 95.6 (3.9) | 95.9 (3.3) | 128.5 (17.4) | 130.3 (16.0) |
| | Approach | 17.4 (6.8) | 17.2 (4.7) | 94.7 (4.9) | 94.9 (4.5) | **95.8 (13.5)** | 100.1 (14.5) |
| A321-111 | Ascent | **3.6 (1.8)** | 5.0 (1.4) | 94.3 (5.5) | **96.5 (5.6)** | **21.7 (2.5)** | 42.4 (18.7) |
| | Climb out | **4.3 (1.7)** | 7.1 (3.1) | **91.7 (9.0)** | 67.6 (22.7) | 21.7 (1.2) | **17.1 (0.9)** |
| | Cruise | **10.5 (5.0)** | 13.0 (5.5) | **98.8 (4.3)** | 89.8 (13.2) | 78.1 (13.6) | **50.5 (3.6)** |
| | Descent | 22.2 (4.7) | 23.4 (5.0) | **90.8 (4.6)** | 88.7 (6.9) | 109.4 (18.5) | 111.4 (12.0) |
| | Approach | 16.4 (4.4) | 17.2 (5.6) | **92.5 (5.4)** | 88.5 (9.6) | **77.6 (16.0)** | 88.9 (10.9) |
| A330-202 | Ascent | **4.5 (2.2)** | 5.7 (1.7) | **96.7 (3.2)** | 93.8 (5.3) | 27.5 (2.6) | **26.0 (1.6)** |
| | Climb out | **4.1 (1.8)** | 8.8 (1.6) | **91.8 (9.7)** | 27.6 (13.8) | 18.3 (1.3) | **9.3 (0.5)** |
| | Cruise | **9.4 (3.8)** | 14.8 (9.9) | 94.4 (16.3) | 98.0 (3.6) | **60.5 (19.3)** | 73.0 (7.3) |
| | Descent | **27.7 (3.8)** | 41.5 (8.0) | **92.1 (4.9)** | 89.1 (5.1) | 176.5 (21.5) | **142.1 (17.9)** |
| | Approach | **30.0 (11.2)** | 38.2 (15.2) | **91.2 (7.5)** | 84.8 (8.4) | 146.3 (26.5) | **115.2 (15.9)** |

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Chapter 6

# Taxi-Out Fuel Burn

In this chapter, the methodologies and insights from the previous chapters are leveraged in order to develop models of surface fuel flow. Taxi-out refers to the portion of ground movement at the departure airport occurring between an aircraft's end of pushback and the start of its takeoff roll. As before, a physical understanding of the taxi-out process is used to assist in model feature selection. The primary modeling objective is to predict the taxi-out fuel flow rate profile for a flight using its taxi-out trajectory information. The fuel flow rates predicted by the models developed in this chapter are compared to those given by AEDT. Only the FDR-II dataset has been used for developing the surface fuel flow rate models. Modeling the taxi-out fuel flow rates is needed to estimate surface fuel burn and emissions generated during taxi.

## 6.1 Model features

A feature engineering approach similar to that adopted in Chapter 5 is used to select the features important for fuel flow modeling in taxi-out. Figure 6-1 shows airplane dynamics during taxi.

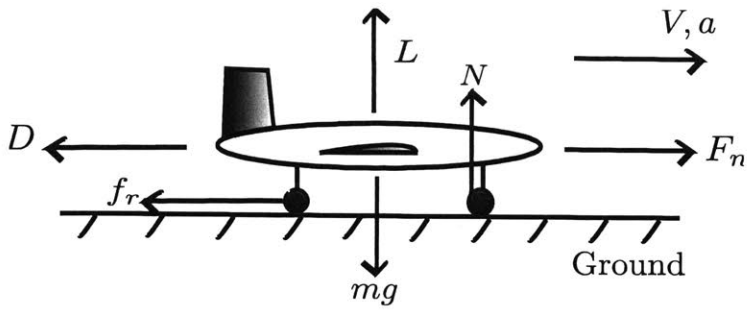



Figure 6-1: Schematic of airplane dynamics during taxi.

The equations of motion during taxi are as follows:

$$L + N = mg \tag{6.1}$$

$$F_n - D - f_r = ma_{\text{taxi}} \tag{6.2}$$

$$L = q\mathcal{S}C_L \tag{6.3}$$

$$D = q\mathcal{S}C_D \tag{6.4}$$

$$f_r = \mu_r N \tag{6.5}$$

$$a_{\text{taxi}} = \frac{dV}{dt} \tag{6.6}$$

$$q = \frac{1}{2}\rho_\infty V^2 \tag{6.7}$$

Here, $N$ is the normal reaction from the ground, $f_r$ is the frictional force from the ground, $a_{\text{taxi}}$ is the aircraft longitudinal acceleration during taxi, and $\mu_r$ is the coefficient of friction.

As done in the previous chapter, the averaged fuel flow rate per engine can be linked to the net thrust via the Thrust Specific Fuel Consumption (TSFC) as follows:

$$\dot{m}_f = \frac{\text{TSFC} \times F_n}{N_{\text{eng}}} \tag{6.8}$$

$$\frac{\text{TSFC}}{\sqrt{\theta_\infty}} = f_{\text{TSFC}}\left(M_\infty, \frac{F_{n,\text{pe}}}{\delta_\infty}, \nu_{\text{eng}}\right) \tag{6.9}$$

The above equations and relations (Equations 6.1–6.9) reveal the following functional dependency for the averaged per engine fuel flow rate in taxi (neglecting constants for a particular

aircraft/engine type):

$$\dot{m}_f\Big|_{\text{aircraft/engine type}} = f_{\dot{m}_f}(\theta_\infty, \delta_\infty, V, m, a_{\text{taxi}}, \mu_r, C_D, C_L, \nu_{\text{eng}}) \tag{6.10}$$

The values of the ambient temperature and pressure on the surface can be obtained from airport weather reports. These values do not tend to vary as rapidly as wind speed and direction. As done in the previous chapter, we consider the dependence of the fuel flow rate on the ground speed (and not the true airspeed). However, since surface wind speeds are typically not high, the aircraft airspeed is likely to be equal to the ground speed ($V \approx V_{\text{GS}}$). In addition, we assume that the mass of the aircraft does not vary appreciably during the course of taxi-out, and that it equals the takeoff mass $m_{\text{TO}}$. The coefficient of friction during taxi (which is a function of time due to braking action by the pilot), the drag and lift coefficients, and the engine component efficiencies cannot be estimated from trajectory data alone, and are not included as model features. The cumulative effect of these unmodeled features is captured through the prediction intervals generated by the statistical models. Therefore, an approximate functional relationship for the fuel flow rate in taxi-out is given by

$$\dot{m}_f\Big|_{\text{aircraft/engine type}} \approx \dot{m}_{f_{\text{taxi}}}(\theta_\infty, \delta_\infty, V_{\text{GS}}, \dot{m}_{\text{TO}}, a_{\text{taxi}}). \tag{6.11}$$

Thus, in the regression models to be built, the input/independent/predictor variables are the following:

- Ambient temperature normalized by the ISA sea level static temperature ($\theta_\infty$, a dimensionless quantity)
- Ambient pressure normalized by the ISA sea level static pressure ($\delta_\infty$, a dimensionless quantity)
- Aircraft ground speed ($V_{\text{GS}}$, in $\text{m}\cdot\text{s}^{-1}$)
- Aircraft takeoff mass ($m_{\text{TO}}$, in kg)
- Aircraft longitudinal acceleration during taxi ($a_{\text{taxi}}$, in $\text{m}\cdot\text{s}^{-2}$)

The output/dependent/predicted variable is the averaged fuel flow rate per engine ($\dot{m}_f$, in kg·s$^{-1}$). It should be noted that the normalized values of the ambient temperature and pressure (instead of the non-normalized values) are used as predictors as the ICAO Databank with Boeing Fuel Flow Method 2 (BFFM2) correction uses these variables to model the fuel flow rate.

## 6.2 Taxi-out fuel flow rate profile

Figure 6-2 shows a typical fuel flow rate profile during taxi-out (as seen in the FDR-II dataset). It can be seen that the fuel flow rate profile (red curve) can be divided into two distinct regions: A baseline region and a fuel flow spike region. The baseline region is characterized by extended intervals of an almost constant (low variation) fuel flow rate having a low value. The fuel flow spike region is characterized by shorter-interval spikes in the fuel flow rate with values generally greater than the baseline fuel flow rate. Therefore, these two fuel flow rate regions need to be modeled separately. Figure 6-2 also shows a mean baseline fuel flow rate (in blue) obtained by averaging the baseline fuel flow rates for a particular taxi-out operation.

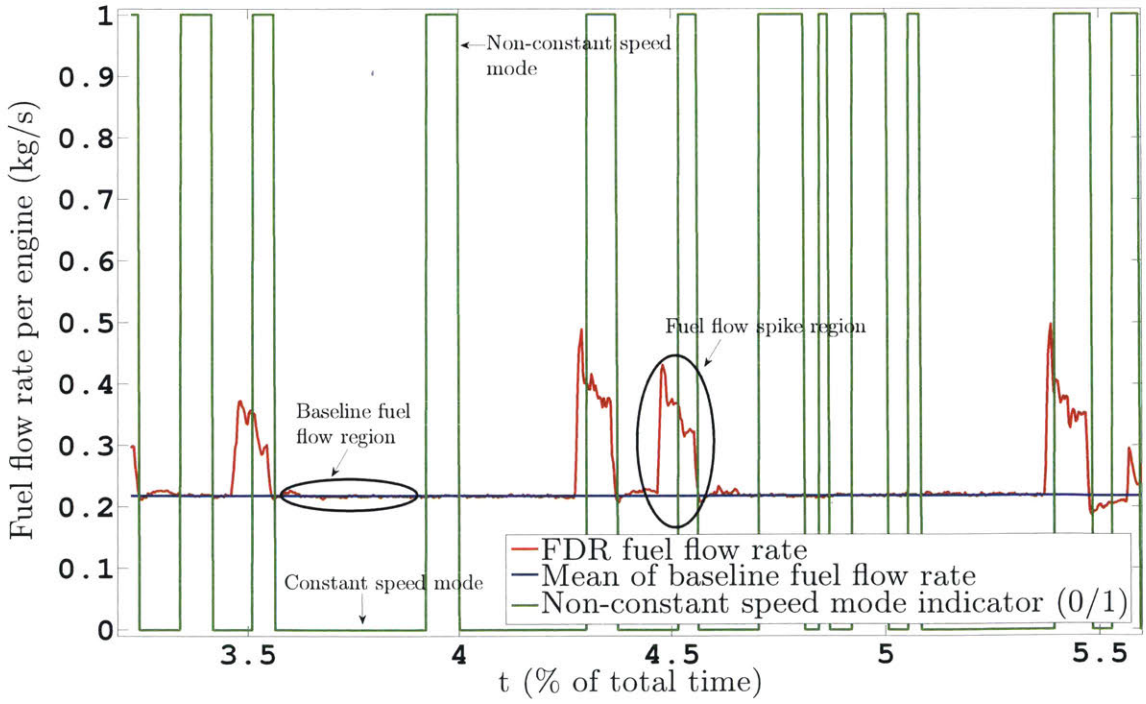



Figure 6-2: Typical fuel flow rate profile in taxi-out.

Table 6.1 shows different characteristics of the baseline fuel flow region for the A330-343 and the B777-300ER aircraft types seen in the FDR-II dataset. It can be seen that on an average, more than 90% of the taxi-out fuel consumption occurs during the baseline fuel flow region. Thus in this thesis, only the baseline fuel flow region is modeled and the fuel flow spikes are neglected.

Table 6.1: Taxi-out: Characteristics of the baseline fuel flow region. The table shows the mean and the range of time spent and fuel mass consumed in the baseline fuel flow region, as a percentage of the total time and fuel burn in taxi-out.

| A/C Type | Time (%) | | Fuel Burn (%) | |
|---|---|---|---|---|
| | Mean | Range | Mean | Range |
| A330-343 | 94.1 | 76.1 – 100.0 | 91.0 | 68.0 – 100.0 |
| B777-300ER | 93.0 | 77.4 – 100.0 | 91.0 | 73.0 – 100.0 |

## 6.3 Trajectory smoothing

The objective of this chapter is to develop models which can map fuel flow rates to aircraft trajectories during taxi-out. However, raw trajectory measurements are often found to be noisy and could have low resolution. Moreover, in Section 6.1 it is shown that acceleration during taxi is a predictor variable for fuel flow rate modeling. The values of aircraft acceleration during taxi are generally not explicitly recorded in the trajectory data. Hence, the raw trajectory data need to be smoothed in order to estimate the variables of interest (such as acceleration) before developing a fuel flow rate model [108].

### 6.3.1 Aircraft kinematics

The aircraft trajectory on the surface is modeled using the following two-dimensional kinematic model:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{B_w}\mathbf{w_n} \tag{6.12}$$

$$\mathbf{x} = \begin{bmatrix} x(t) \\ y(t) \\ V_{\mathrm{GS}}(t) \\ \theta_c(t) \end{bmatrix} \; ; \; \mathbf{f}(\mathbf{x}) = \begin{bmatrix} V_{\mathrm{GS}} \sin(\theta_c(t)) \\ V_{\mathrm{GS}} \cos(\theta_c(t)) \\ a_{\mathrm{taxi}}(t) \\ \omega(t) \end{bmatrix} \; ; \; \mathbf{B_w} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \; ; \; \mathbf{w_n} = \begin{bmatrix} w_a \\ w_\theta \end{bmatrix} \tag{6.13}$$

Here, $\mathbf{x}$ is the state vector be estimated, $(x, y)$ is the aircraft position on the ground (as a function of time), $V_{\mathrm{GS}}$ is the ground speed, $\theta_c$ is the course track angle, $a_{\mathrm{taxi}}$ is the acceleration, $\omega$ is the turn rate and $\mathbf{w_n} \sim \mathcal{N}(0, \mathbf{W})$ is the process noise with zero mean and covariance $\mathbf{W}$. The process noise accounts for the uncertainty in the acceleration and turn rate. Here, $w_a$ and $w_\theta$ are the components of the process noise vector and are assumed to be uncorrelated and independent. The trajectory measurements provide full information about the state vector ($\mathbf{x}$) that needs to be estimated. It should be noted that the actual measurements contain latitude and longitude coordinates, which are converted into local $(x, y)$ coordinates. The measurement vector, $\mathbf{z}$ is given by

$$\mathbf{z} = \mathbf{x} + \mathbf{v_n}. \tag{6.14}$$

Here, $\mathbf{v_n} \sim \mathcal{N}(0, \mathbf{R_n})$ is the additive noise (with zero mean and covariance $\mathbf{R_n}$), assumed to be independent and uncorrelated with the process noise.

### 6.3.2 Smoothing algorithm

The state estimation is done using a fixed interval smoother since measurements exist for the entire trajectory [109]. We are interested in computing $p(\mathbf{x}_t|\mathbf{z}_{1:t_{\mathrm{tot}}})$ (the distribution of the state vector at time $t$ given the complete measurement vector ending at time $t_{\mathrm{tot}}$), unlike filtering whose output is $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ (which only takes measurements up to time $t$ into account). For estimating the state variables, the aircraft kinematics is approximated by a hybrid system of the

following form:

$$\dot{\mathbf{x}} = \mathbf{f}^k(\mathbf{x}) + \mathbf{B_w}\mathbf{w_n};\ \ \mathbf{f}^k(\mathbf{x}) = \begin{bmatrix} V_{\mathrm{GS}} \sin(\theta_c(t)) \\ V_{\mathrm{GS}} \cos(\theta_c(t)) \\ a_{\mathrm{taxi}_k} \\ \omega_k \end{bmatrix};\ k \in 1, 2, \ldots, \mathcal{M} \tag{6.15}$$

The models are specified by $\mathcal{M}$ modes of operation, each of which is a set $(a_{\mathrm{taxi}_k}, \omega_k)$. The hybrid state vector consists of the original state variables ($\mathbf{x}$) as well as the mode of operation. Multi-model kinematics provides better tracking accuracy compared to single model kinematics. Another advantage of using multi-model kinematics is that it helps to identify the 'most likely mode' of the system at any instant. The estimation is performed on the discretized system with time discretization equal to the sampling rate of the measurements in the FDR-II dataset (1 Hz). The algorithm is based on combining estimates produced by an Interacting Multiple Model (IMM) filter with a backward-time recursion [110, 108]. The algorithm gives out smoothed estimates of the state variables (trajectory) as well as the probability of the system being in a particular mode (i.e., the mode fraction). The 'most likely mode' of the system at any time instant is determined from the maximum a posteriori probability, i.e., the mode with the greatest mode fraction. Table 6.2 shows the modes considered for analyzing the taxi trajectories.

Table 6.2: Different modes considered in taxi trajectory smoothing.

| **Mode Number** | **Mode Name** | $a_{\mathrm{taxi}}$ **Threshold** ($\mathbf{m \cdot s^{-2}}$) | $\omega$ **Threshold** ($\mathbf{deg \cdot s^{-1}}$) |
|---|---|---|---|
| 1 | Constant speed | 0 | 0 |
| 2 | Acceleration | 0.35 | 0 |
| 3 | Deceleration | -0.35 | 0 |

The modes correspond to constant speed, acceleration and deceleration events. The modes are motivated by the fact that high acceleration may contribute to fuel burn spikes, whereas, turns may not have any significant impact on the fuel flow. Another reason for not including turns is that accelerated turns might show up as turns and not acceleration events.

Figure 6-3 shows the comparison of the flight trajectory between the FDR data and the smoothed estimates for a departing flight on the surface. It can be seen that the estimated

trajectory is smooth as compared to the raw FDR data. The zig-zag pattern of the raw data is due to poor resolution. Figure 6-4 shows a comparison of the ground speed profile for the flight trajectory between the FDR data and the smoothed estimate. Even the ground speed measurements from the FDR data have a bad resolution, which is smoothed out in the estimates.

(a) Complete trajectory (b) Zoomed-in image

Figure 6-3: Comparison of taxi-out trajectory on the airport surface between the FDR raw data (blue) and smoothed estimates (red) for one flight at the Boston Logan International airport (BOS).

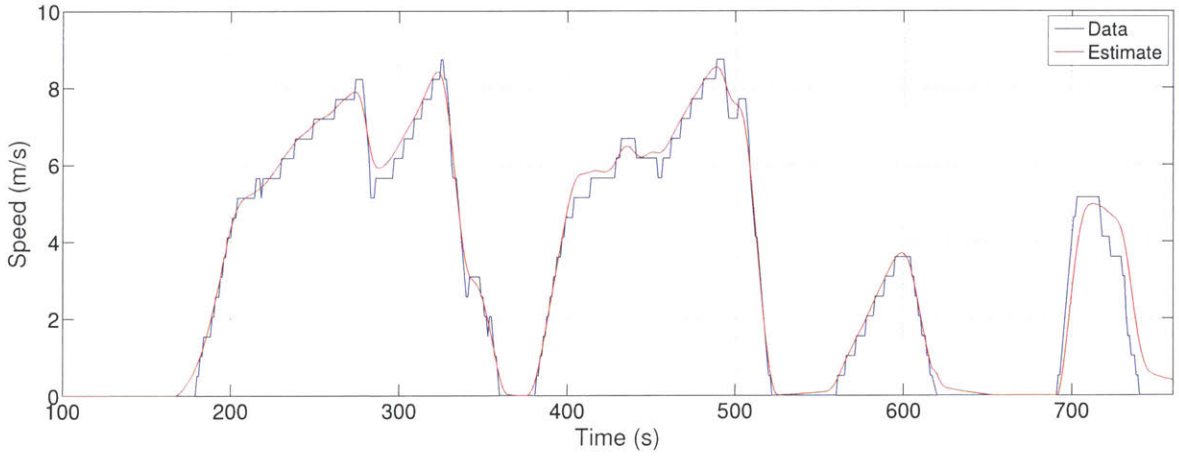



Figure 6-4: Comparison of taxi-out ground speed profile between FDR raw data and smoothed estimates.

### 6.3.3 Acceleration estimates

One of the objectives of the smoother is to derive smoothed estimates of acceleration during taxi. Figure 6-2 shows the non-constant speed mode indicator variable (in green). Parts of the trajectory identified to have the acceleration/deceleration (i.e., non-constant speed) mode as the 'most likely mode' are given the indicator 1. The other parts of the trajectory (having a constant speed mode) are given the indicator 0. Analysis of the fuel flow rate and acceleration profiles for different flights in the FDR-II dataset reveals a strong correlation between the fuel flow spike regions and the regions having an acceleration mode. This is expected as pushing the throttle forward results in a fuel flow spike that causes the aircraft to accelerate. On the other hand, there is a good correlation between the baseline fuel flow regions and the regions having a constant speed mode. However, cases have been observed in the data when an acceleration mode is observed without any accompanying fuel flow spike (false positives). This can happen when the pilot releases the brakes at an unchanged throttle setting causing the aircraft to accelerate. Cases have also been observed when a fuel flow spike corresponds to a non-acceleration mode (false negatives). This can happen when pushing the throttle forward is accompanied by the pilot pressing the brakes in a manner so as to not cause the aircraft to accelerate. False negatives can also occur due to the acceleration being small compared to the threshold, resulting in it being flagged as a constant speed event.

The smoother output depends significantly on the value of the acceleration threshold. The acceleration threshold is chosen such that the false positives and false negatives are almost the same. Analysis of different acceleration thresholds shows that an acceleration threshold of 0.35 $ms^{-2}$ balances the false positives and false negatives for both the A330-343 and the B777-300ER flights in the FDR-II dataset. Thus, this value of the acceleration threshold is chosen for taxi fuel flow rate modeling for all the flights in the FDR-II dataset. This acceleration threshold is also recorded in Table 6.2.

The estimates of the mode fractions given by the smoother can be used to estimate the taxi acceleration by the following equation:

$$\hat{a}_{\text{taxi}} = a_{\text{thres.}}(f_2 - f_3) \tag{6.16}$$

Here, $\hat{a}_{\text{taxi}}$ is the estimate of the taxi acceleration from the smoother, $a_{\text{thres.}}$ is the acceleration threshold employed, and $f_2$ and $f_3$ are the mode fractions corresponding to the acceleration mode and the deceleration mode, respectively.

## 6.4 Baseline fuel flow modeling

As mentioned in Section 6.3.3, the baseline fuel flow rate regions correlate well with the constant speed mode portions of the taxi-out trajectory. Thus, in this thesis, the averaged baseline fuel flow rate per engine is modeled using the predictor variables from only the constant speed mode portions of the taxi trajectory. Smoothed estimates of these predictors are used for model building. For each aircraft type in the FDR-II dataset, 65% of all the flights constitute the training set, 15% constitute the validation set, and 20% constitute the test set. Table 6.3 shows the number of flights and the number of observations corresponding to baseline fuel flow and constant speed mode for each of the training, validation, and test datasets used in this chapter. Using the training set, two models are built to predict the baseline fuel flow rate. These models are hereafter, referred to as the Baseline-1 and Baseline-2 models.

Table 6.3: Details of training, validation, and test datasets drawn from the FDR-II dataset with the number of flights (# Flts.) and the number of observations (# Obs.) for each aircraft type in the taxi-out phase.

| **A/C Type** | **Training Set** | | **Validation Set** | | **Test Set** | |
|---|---|---|---|---|---|---|
| | # Flts. | # Obs. | # Flts. | # Obs. | # Flts. | # Obs. |
| A330-343 | 117 | 70,785 | 27 | 15,962 | 37 | 24,117 |
| B777-300ER | 81 | 41,643 | 19 | 7,954 | 25 | 10,669 |

### 6.4.1 Baseline model, version 1

The Baseline-1 model is a simple model which models the baseline fuel flow rate as a constant for a particular taxi-out operation. It therefore, neglects the variation seen in the baseline fuel flow rate within the same taxi-out operation. The aim of developing this simple model is to suggest an improvement over the ICAO Databank - BFFM2 (ICAO-BFFM2) method, used in

AEDT to estimate fuel flow rate during taxi. The ICAO Databank method with BFFM2 correction (ICAO-BFFM2) estimates the fuel flow rate per engine during taxi through the following equation (the Mach number dependence from the original equation has been dropped due to very low Mach numbers during taxi):

$$\dot{m}_{f_{\text{ICAO-BFFM2,taxi}}} = 1.1\dot{m}_{f_{\text{ICAO,taxi}}}\delta_\infty\theta_\infty^{-3.8} \tag{6.17}$$

$$\approx 1.1\dot{m}_{f_{\text{ICAO,taxi}}} \tag{6.18}$$

Here, $\dot{m}_{f_{\text{ICAO,taxi}}}$ is the ground idle taxi fuel flow rate enumerated in the ICAO Databank. The approximation is reasonable as $\delta_\infty$ and $\theta_\infty$ are close to 1 on the ground. This approximation (Equation 6.18) is used in AEDT to model taxi fuel flow rate per engine and is used in the remainder of this chapter as the example of the current state-of-the-practice model to estimate taxi fuel flow.

For each taxi-out operation in the training set, the mean baseline fuel flow rate per engine in taxi-out (blue curve in Figure 6-2) is regressed against the mean of the values of the corrected ambient temperature ($\theta_\infty$) and the corrected ambient pressure ($\delta_\infty$) in the constant speed mode portions of the same taxi-out operation. Only these two predictor variables are chosen out of the complete set of predictor variables mentioned in Section 6.1 in accordance with the ICAO-BFFM2 fuel flow rate modeling relation in Equation 6.17 (as the purpose of the Baseline-1 model is to improve upon the ICAO-BFFM2 method). An Ordinary Least Squares (OLS) regression approach is found to be sufficient to develop this simple model. (The sufficiency of OLS regression is verified by conducting diagnostic tests which show that the assumptions of OLS regression, as explained in Section 4.2, are satisfied. The regression coefficients are also found to be statistically significant at a 5% significance level.) Table 6.4 shows the OLS regression-derived equations for modeling the fuel flow rate using the Baseline-1 model.

Table 6.4: Baseline-1 model equations to model fuel flow rate per engine during taxi-out.

| A/C Type | Baseline-1 Model Equation |
|---|---|
| A330-343 | $\dot{m}_{f_{\text{taxi}}} = 0.779\dot{m}_{f_{\text{ICAO,taxi}}}\delta_\infty\theta_\infty^{0.350}$ |
| B777-300ER | $\dot{m}_{f_{\text{taxi}}} = 0.753\dot{m}_{f_{\text{ICAO,taxi}}}\delta_\infty\theta_\infty^{0.717}$ |

The main difference between the AEDT equation and the Baseline-1 model equations is the change in the leading factor from 1.1 to 0.779 for the A330-343 and 0.753 for the B777-300ER.

### 6.4.2 Baseline model, version 2

In contrast with the Baseline-1 model, Baseline-2 model does not model the baseline fuel flow rate as a constant. This more complicated model can therefore, capture variation in the baseline fuel flow rate occurring during the same taxi-out operation. For each taxi-out operation in the training set, the instantaneous baseline fuel flow rate per engine in taxi-out (red curve in Figure 6-2 corresponding to baseline fuel flow) is regressed against the instantaneous values of all the predictor variables selected in Section 6.1. Only those instants in the taxi-out operation are used for model building which correspond to baseline fuel flow as well as the constant speed mode. Based on insights gained in the previous chapter, Gaussian Process Regression (GPR) is used to train the models. Analysis of model predictive performance on the validation sets leads to the Dot Product Exponential (DPE) kernel being selected for Baseline-2 model building.

## 6.5 Performance of fuel flow rate model

Table 6.5 shows the predictive performance of the two baseline fuel flow rate models on unseen test data not used for training or validation. The predictive performance of each model is evaluated in two ways:

1. Evaluation using test set outputs similar to training set outputs **(Case #1)**: In this way of evaluation, the models are evaluated by comparing their predictions with the same output variables in the test set as those used for model training. Thus, Baseline-1 model is evaluated by comparing its predictions with the mean baseline fuel flow rates in the test set. The test inputs are the mean values of the appropriate predictors in the constant speed mode regions, similar to the way the model is trained. Baseline-2 model is evaluated by comparing its predictions with the instantaneous baseline fuel flow rates in the test set. The test inputs are the instantaneous values of the appropriate predictors in the constant

speed mode and baseline fuel flow regions, similar to the way the model is trained. This type of analysis is done to evaluate the model quality.

2. Evaluation using the complete taxi-out test trajectory (**Case #2**): In this way of evaluation, both the models are evaluated by using them to predict the fuel flow rate at each instant of the taxi-out trajectory in the test set, including the instants corresponding to the fuel flow spike or the non-constant speed mode regions (even though the models have not been trained for such portions). This way of evaluation mimics the way in which the developed models will be practically used to predict taxi fuel flow rates, given the complete trajectory for a particular taxi-out operation.

Table 6.5 also shows the predictive performance of the AEDT models (Equation 6.18) on the test data. The AEDT models do not give prediction intervals for the fuel flow rates and hence, the Prediction Coverage (PC) and the Normalized Length of Prediction Intervals (NLPI) for these models are not reported in the table. The Mean Errors (MEs) of the data-driven models are statistically compared (Section 4.6) with zero to check if the models are biased. The Mean Absolute Errors (MAEs) of the data-driven models and the AEDT models are also statistically compared among one another to determine which model gives the lowest MAE. PC for the two baseline models are statistically compared to one another to determine which model gives the higher PC. Bold entries in the table indicate statistically significant results (models with bias, the lowest MAE, the highest PC) at the 5% significance level.

Case #1 analysis shows that both the baseline fuel flow rate models give PCs that are close to 95%, indicating good model quality. When the baseline models are evaluated on the complete taxi trajectory (Case #2), the mean predictions for both the models are found to be biased. For both the aircraft types, the MAEs given by the two baseline models are statistically similar. However, the MAEs for the data-driven models developed are found to be statistically significantly lower than those given by the AEDT models. For the A330-343, the MAE for both the baseline models is 6.3%. This represents a reduction in MAE of about 84% as compared to the AEDT model. For the B777-300ER, the MAE for both the baseline models is 2.7%. This represents a reduction in MAE of about 94% as compared to the AEDT model. Unlike the AEDT models, the data-driven models can also quantify uncertainty in the fuel flow rates. The

Table 6.5: Performance of the different data-driven baseline fuel flow rate models and the AEDT models to predict fuel flow rates on unseen test data during taxi-out. Each cell reports the mean and standard deviation (within parentheses) of the evaluation metric across all the flights. 'Bas.1' and 'Bas.2' refer to the Baseline-1 and the Baseline-2 models, respectively.

| A/C Type | Evaluation Method | ME (%) | | | MAE (%) | | | PC (%) | | NLPI (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bas.1 | Bas.2 | AEDT | Bas.1 | Bas.2 | AEDT | Bas.1 | Bas.2 | Bas.1 | Bas.2 |
| A330-343 | Case #1 | -0.8 | **-0.4** | **39.3** | **2.7** | **4.0** | 40.9 | 94.6 | 91.6 | 8.5 | 14.9 |
| | | (7.9) | **(8.5)** | **(11.3)** | **(7.5)** | **(7.8)** | (4.4) | (22.9) | (16.4) | (0.0) | (2.4) |
| | Case #2 | **-3.3** | **-2.9** | **36.0** | **6.3** | **6.3** | 39.4 | 71.5 | **84.8** | 8.5 | 15.1 |
| | | **(7.4)** | **(7.3)** | **(10.6)** | **(6.9)** | **(6.7)** | (4.4) | (21.3) | **(16.4)** | (0.0) | (2.3) |
| B777-300ER | Case #1 | -0.0 | 0.0 | **45.2** | **0.6** | **1.0** | 45.2 | 92.0 | 95.8 | 2.6 | 6.7 |
| | | (0.8) | (0.8) | **(1.0)** | **(0.4)** | **(0.4)** | (1.0) | (27.7) | (3.3) | (0.0) | (0.2) |
| | Case #2 | **-1.8** | **-1.8** | **42.6** | **2.7** | **2.7** | 43.2 | 72.9 | **89.7** | 2.6 | 6.8 |
| | | **(1.1)** | **(1.1)** | **(1.6)** | **(1.2)** | **(1.2)** | (1.5) | (23.6) | **(6.8)** | (0.0) | (0.2) |

Baseline-2 models give a statistically significantly larger PC as compared to that given by the Baseline-1 models. This increased coverage results from the ability of the Baseline-2 model to incorporate variability in the baseline fuel flow rate. The Baseline-2 model gives a PC of 84.8% for the A330-343 and 89.7% for the B777-300ER. However, this increased coverage is the result of a larger NLPI for the Baseline-2 model. It should be noted that Case # 2 indicates a degradation in model predictive performance as compared to Case #1. This degradation is expected as in Case #2, the models are being used to predict the fuel flow rates for the entire taxi-out trajectory, including the fuel flow spike or the non-constant speed mode regions for which the baseline models have not been trained.

Figures 6-5 and 6-6 show the taxi-out fuel flow rate profiles (predicted as well as the FDR ground truth) for one test data flight each of the A330-343 and the B777-300ER, respectively. It can be observed how the bigger prediction intervals given by the Baseline-2 models account for greater variability in the baseline fuel flow rate. Both the baseline models give a better predictive performance as compared to the AEDT models.

## 6.6 Taxi-out fuel burn prediction

The predicted taxi-out fuel flow rate can be used for predicting the total mass of fuel consumed during taxi (i.e., the fuel burn). Equations 5.26 – 5.28 are similarly applied to predict the taxi-out

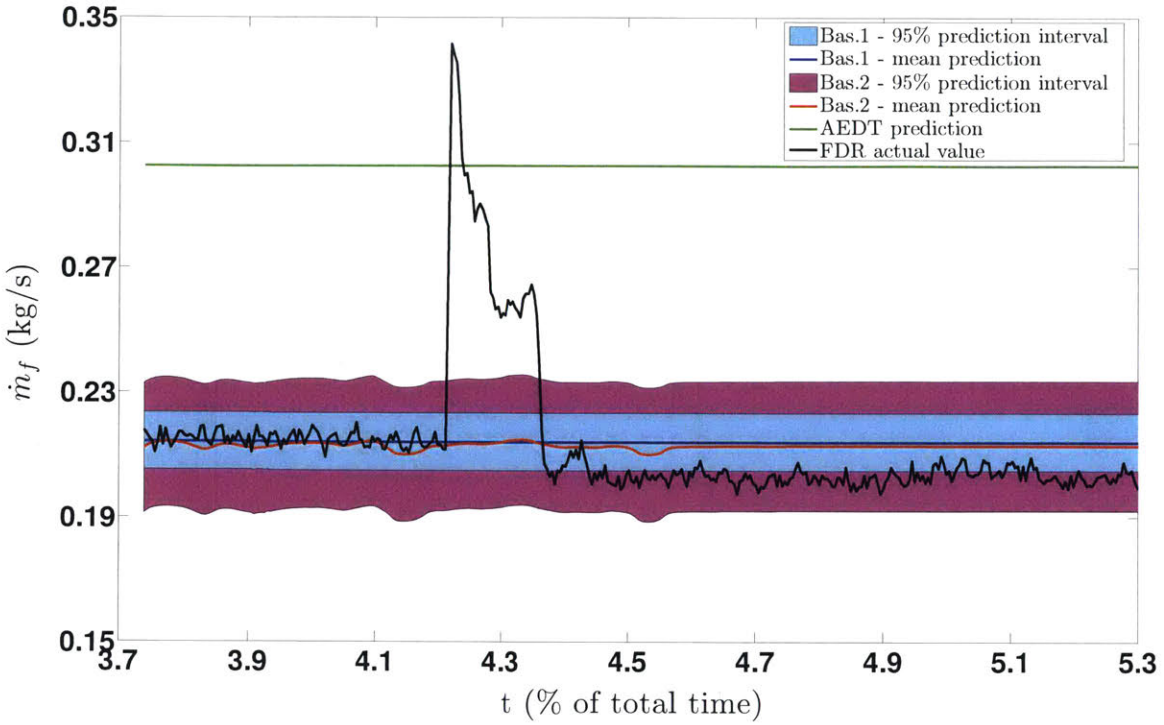



Figure 6-5: A330-343: Fuel flow rate prediction for one test data flight in taxi-out. The x-axis represents time since the start of the FDR record as a percentage of the total flight time.

fuel burn.

Table 6.6 shows the fuel burn predictive performance using the Baseline-1, Baseline-2, and the AEDT models on the unseen prediction set not used for model training (i.e., the combined validation and test sets). The PC and NLPI are only reported for the Baseline-2 models. This is because as mentioned in Section 6.5, the Baseline-1 model is restricted to estimating the mean baseline fuel flow rate (and not the uncertainty therein) and the AEDT model being deterministic does not give prediction intervals. For the numbers in the table, the models are run on the complete taxi-out trajectories, including regions of fuel flow spikes or non-constant speed modes (even though the underlying fuel flow rate models have not been trained for these regions). Statistical multi-comparison tests are performed to compare the MAEs of the different models among one another and to compare the MEs of the different models with zero. Bold entries in the table indicate statistically significant results (models with bias, the lowest MAE) at the 5% significance level. From Table 6.6, it can be observed that for both the aircraft types in

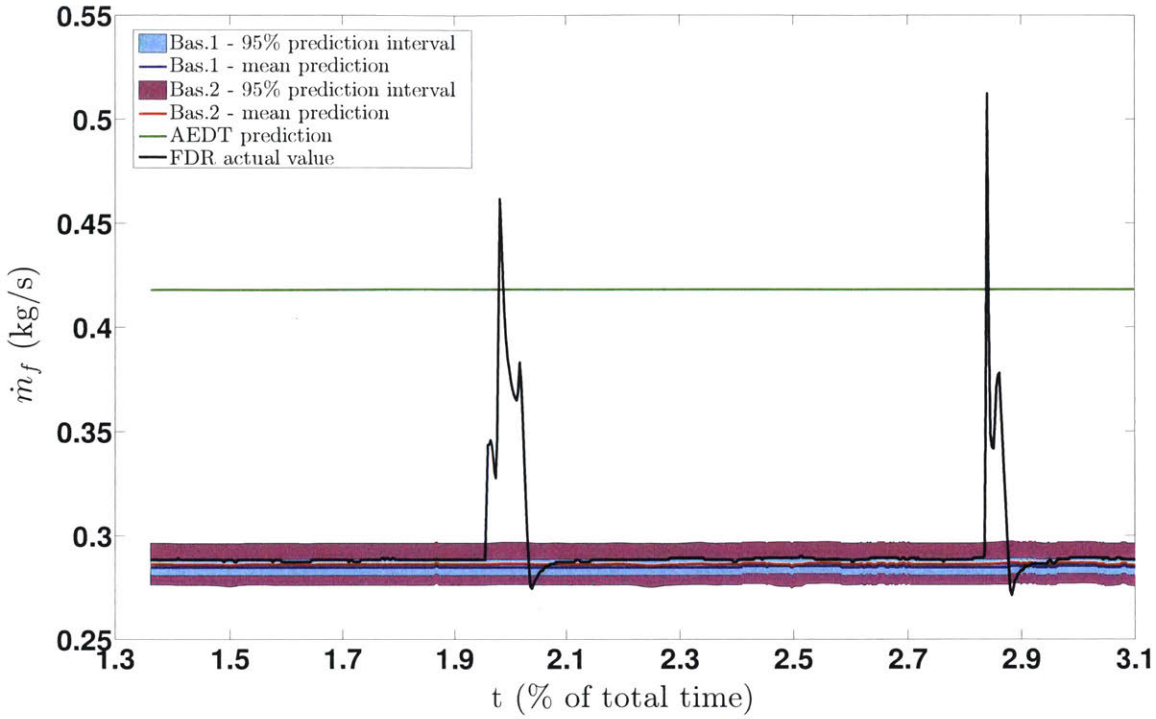



Figure 6-6: B777-300ER: Fuel flow rate prediction for one test data flight in taxi-out. The x-axis represents time since the start of the FDR record as a percentage of the total flight time.

Table 6.6: Performance of the different data-driven baseline fuel flow rate models and the AEDT models to predict fuel burn on unseen prediction dataset not used for training (combined validation and test sets). Each cell reports the mean and standard deviation (within parentheses) of the evaluation metric across all the flights. 'Bas.1' and 'Bas.2' refer to the Baseline-1 and the Baseline-2 models, respectively.

| A/C Type | ME (%) | | | MAE (%) | | | PC (%) | NLPI (%) |
|---|---|---|---|---|---|---|---|---|
| | **Bas.1** | **Bas.2** | **AEDT** | **Bas.1** | **Bas.2** | **AEDT** | **Bas.2** | **Bas.2** |
| A330-343 | **-4.7 (5.9)** | **-4.2 (6.0)** | **31.2 (19.8)** | **4.9 (5.7)** | **4.6 (5.7)** | 31.8 (18.7) | 87.5 (33.3) | 15.0 (1.3) |
| B777-300ER | **-3.0 (1.9)** | **-3.0 (1.9)** | **41.1 (2.8)** | **3.0 (1.8)** | **3.0 (1.8)** | 41.1 (2.8) | 65.9 (47.9) | 6.9 (0.2) |

the FDR-II dataset, the baseline fuel flow rate models give statistically significantly biased predictions of the fuel burn. Both the models give more accurate mean predictions of the fuel burn as compared to the AEDT method. The two baseline fuel flow rate models do not statistically significantly differ from each other in terms of the MAE in fuel burn prediction. Thus, both the models give similar mean predictions of the fuel burn. For the A330-343, the MAE for the

Baseline-1 model is 4.9% and that for the Baseline-2 model is 4.6%. Thus, these baseline models can achieve a reduction in MAE of as much as about 86% for the A330-343 over its AEDT model. For the B777-300ER, the MAE for both the baseline models is 3.0%. Thus, the baseline models can achieve a reduction in MAE of as much as about 93% for the B777-300ER over its AEDT model. Despite being trained on only baseline fuel flow rates, the Baseline-2 model gives 87.5% PC for the A330-343 fuel burn predictions and 65.9% PC for the B777-300ER fuel burn predictions on the complete taxi-out trajectory.

Figures 6-7 and 6-8 contain box-plots showing the errors and the absolute errors, respectively, for the taxi fuel burn prediction across the different flights in unseen data not used for training. Model predictions are unbiased (at a 5% significance level) if the error box-plots in Figure 6-7 contain zero in their notched regions. The notched regions of the absolute error box-plots in Figure 6-8 for the baseline models can be compared to those of the box-plots for the corresponding AEDT models to determine which model gives a statistically significantly lower absolute error (at a 5% significance level) on the taxi fuel burn prediction. Figure 6-8 shows that the absolute errors in fuel burn prediction given by the data-driven models developed are statistically significantly lower than those given by the AEDT models.

It should be noted that the predictive performance of the models developed can be further improved by modeling the fuel flow spikes too. Fuel burn predictions can be improved by explicitly modeling the fuel mass consumed (instead of predicting it via the fuel flow rates). These approaches are not pursued further in this thesis.

## 6.7 ASDE-X data as inputs

In the previous sections, the fuel flow rate and fuel burn model predictive performance was evaluated using unseen data drawing values of the predictor variables from the FDR-II dataset. In practice however, the models will be applied to trajectory data obtained through ground surveillance systems. ASDE-X, described in Chapter 3, is one such surveillance system. In this section, model predictive performance is evaluated using values of the trajectory-based predictor variables (ground speed and taxi acceleration) drawn/estimated from the ASDE-X dataset.

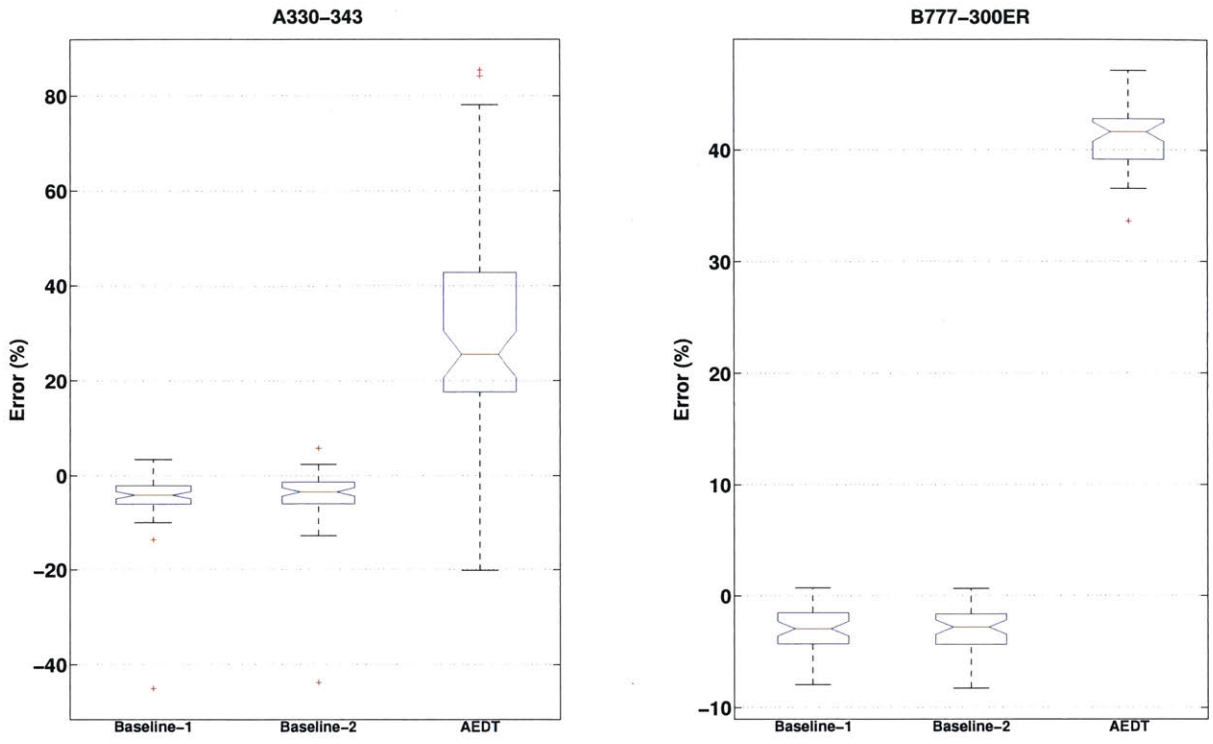



Figure 6-7: Box-plots showing error on taxi-out fuel burn prediction for flights in unseen data not used for training. These box-plots are helpful in checking for model bias in the mean predictions.

The values of the ambient conditions and TOW are still obtained from the FDR-II dataset. The ground truth value of the fuel flow rate needed for model evaluation is also obtained from the FDR-II dataset. The values of the evaluation metrics determined in this section indicate the predictive performance of the models as they will be used in practice using ground surveillance data.

### 6.7.1 ASDE-X trajectory smoothing

Fuel flow rate prediction using ASDE-X data is more challenging because of two reasons: (a) ASDE-X data are more noisy when compared to FDR data, and (b) ASDE-X data have a lot of missing fields [108]. The same modes that were used for smoothing the FDR data (Section 6.3) are used for smoothing the ASDE-X data. The noise parameters are tuned appropriately and the acceleration threshold is chosen to be slightly higher when compared to the FDR case

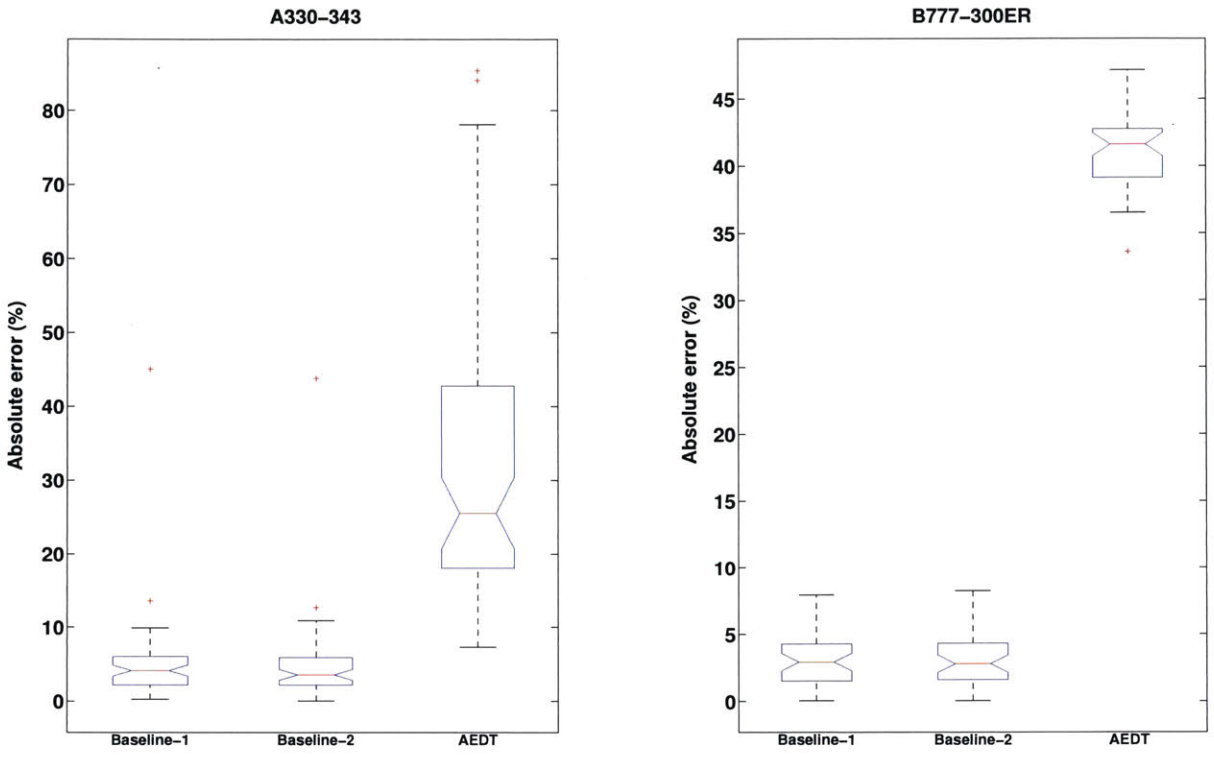



Figure 6-8: Box-plots showing absolute error on taxi-out fuel burn prediction for flights in unseen data not used for training. These box-plots can be used to compare the baseline models with the corresponding AEDT models for the accuracy of the mean predictions.

(0.5 $ms^{-2}$ instead of 0.35 $ms^{-2}$). The threshold is determined by balancing the false positives and false negatives of the acceleration events in the ASDE-X records. Here, false positives refer to the acceleration events (i.e., when a time instant has the acceleration mode) that show up in the ASDE-X estimates but not in the FDR estimates. Similarly, false negatives refer to the acceleration events that do not show up in the ASDE-X estimates but are present in the FDR estimates. Figure 6-9 shows the indicator variable for the acceleration mode obtained from the estimation of the ASDE-X data and the FDR data. It can be seen that the first acceleration event estimated using the FDR data is not seen in the ASDE-X estimates. This is an example of a false negative. Figure 6-10 shows the comparison of the ground speed profile for a flight trajectory obtained from the smoothed ASDE-X estimates with the raw ASDE-X data and the smoothed FDR estimates. It can be seen that the raw ASDE-X measurements are noisy and missing for a good fraction of the time. However, the smoothed ASDE-X estimates compare well with the

smoothed FDR estimates.

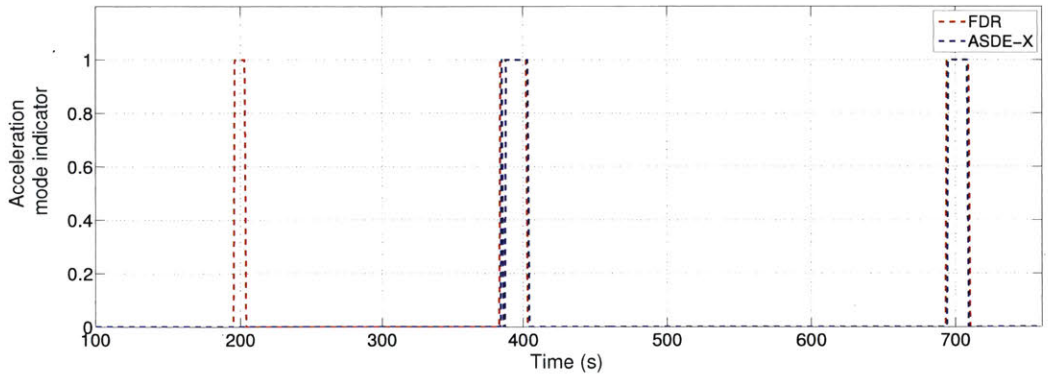



Figure 6-9: Detection of the acceleration mode estimated from ASDE-X and FDR data.

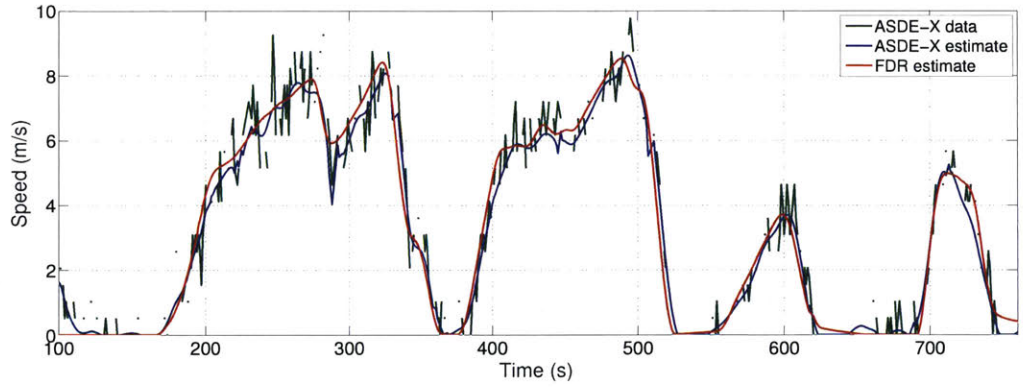



Figure 6-10: Comparison of the ground speed profile for a taxi-out trajectory among raw ASDE-X data, smoothed ASDE-X estimates, and smoothed FDR estimates.

### 6.7.2 Predictive performance with ASDE-X data inputs

Models trained on FDR data in Section 6.4 are applied to trajectory predictor variable (i.e., ground speed and taxi acceleration) values drawn from the ASDE-X data to determine mean predictions and prediction intervals for the fuel flow rates during taxi. Values of the ambient temperature, ambient pressure, TOW, and the ground truth fuel flow rate are still obtained from the corresponding FDR records. Model evaluation is done on 36 flights for the A330-343 and 23 flights for the B777-300ER which have records in both the FDR-II and the ASDE-X datasets and which have not been used for model training.

Table 6.7 tabulates the fuel flow rate predictive performance on the complete taxi trajectory when the source of trajectory variables is ASDE-X. For reference, the predictive performance when the trajectory variables are obtained from FDR data is also tabulated. It should be noted that this predictive performance on FDR-obtained variables may differ from that recorded in Table 6.5 as different sets of flights have been used for model evaluation in the two tables. Bold entries in the table indicate statistically significant results (models with bias, the lowest MAE, the highest PC) at the 5% significance level.

Table 6.7: Performance of the different data-driven baseline fuel flow rate models and the AEDT models to predict fuel flow rates during taxi-out when the trajectory predictor variables are obtained from ASDE-X data. Each cell reports the mean and standard deviation (in parentheses) of the evaluation metric across all the flights. For reference, model predictive performance when the trajectory predictor variables are obtained from FDR data is also given. 'Bas.1' and 'Bas.2' refer to the Baseline-1 and the Baseline-2 models, respectively.

| A/C Type | Trajectory Source | ME (%) | | | MAE (%) | | | PC (%) | | NLPI (%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Bas.1 | Bas.2 | AEDT | Bas.1 | Bas.2 | AEDT | Bas.1 | Bas.2 | Bas.1 | Bas.2 |
| A330-343 | ASDE-X | **-2.9** (2.5) | **-2.5** (2.7) | **36.4** (3.9) | **5.6** (2.0) | **5.8** (1.9) | 39.1 (3.4) | 71.7 (16.6) | **83.3** (10.4) | 8.5 (0.1) | 14.9 (2.0) |
| | FDR | | **-2.6** (2.8) | | | **5.8** (1.9) | | | **83.3** (10.6) | | 14.8 (2.0) |
| B777-300ER | ASDE-X | **-2.5** (1.3) | **-2.5** (1.4) | **41.7** (2.1) | **3.5** (1.5) | **3.6** (1.4) | 43.4 (2.1) | 73.9 (23.9) | **87.4** (6.9) | 2.6 (0.0) | 6.8 (0.2) |
| | FDR | | **-2.5** (1.5) | | | **3.6** (1.4) | | | **87.4** (6.9) | | 6.8 (0.3) |

Table 6.7 shows that model predictive performance when trajectory variables are obtained from ASDE-X is similar to that when trajectory variables are obtained from FDR. The data-driven models give a better predictive performance than the AEDT models. It should be noted that the Baseline-1 and AEDT models do not need any trajectory variables as inputs and are therefore, independent of the source of such data.

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Chapter 7

# Takeoff Weight

In Chapter 5, the aircraft gross mass at takeoff (operational known as the Takeoff Weight TOW) is identified as an input variable to predict the fuel flow rate. However, the TOW for a particular flight cannot be obtained from readily available data sources as it is considered a competitive parameter by airlines. In this chapter, an algorithm is developed to predict the TOW for a particular flight using information from its takeoff ground roll segment. As done in the previous chapters, the chapter starts with explaining the features deemed important for modeling TOW. Gaussian Process Regression is used to build the models for TOW prediction. The model predictive performance is compared to that of the Aircraft Noise and Performance (ANP) model. The ANP model is a part of AEDT and estimates the TOW as a function of the trip length. The chapter ends with a section on how the TOW prediction can be combined with fuel flow rate prediction. It is worth noting that as per the convention in aviation, the term *weight* in this chapter refers to the *mass* of the aircraft in the physical sense.

## 7.1 Sensitivity of fuel flow rate to takeoff weight

The TOW, which is a predictor variable for the fuel flow rate models, needs to be estimated as its actual value for a particular flight is not available. The accuracy of the fuel flow rate estimation therefore, depends on the accuracy of the TOW estimation. For the fuel flow rate sensitivity analysis, the values of the TOW in the unseen prediction dataset not used for training

(combined validation and test datasets) for each aircraft type are changed systematically by a fixed percentage of the true value to give a modified prediction dataset [80]. All the other predictors in the prediction set are held at their original values (which is an approximation as the other variables might depend on TOW too). GPR models trained using the true TOW in the different airborne phases of flight (Section 5.3.1) are run on the modified prediction dataset, and the mean absolute error in the predicted fuel flow rate is calculated. The variation of this error with the percentage deviation of the TOW in the modified prediction dataset from its true value indicates the sensitivity of the model fuel flow rate predictions to the TOW.

Figure 7-1 shows an example of how the mean absolute error in the predicted fuel flow rate can vary with the deviation in the estimated TOW from its actual value. The figure plots the MAE in the fuel flow rate on the modified prediction dataset from the GPR model developed for the A340-541 in ascent, as a function of the deviation of the estimated TOW from the actual TOW. A positive deviation means that the estimated TOW is greater than the actual TOW, while a negative deviation implies that the estimated TOW is less than the actual TOW. From Figure 7-1, it can be seen that the MAE increases for large magnitudes of deviation from the true TOW.

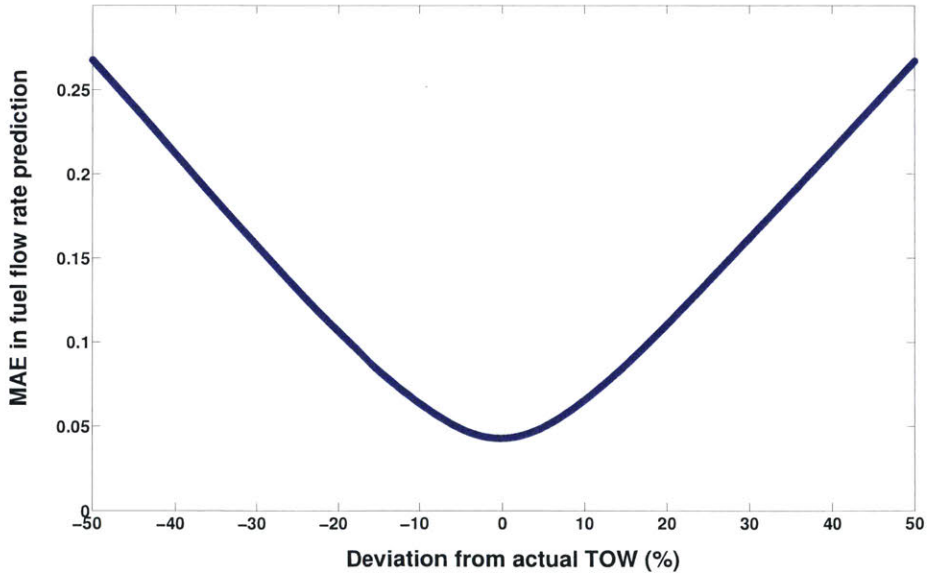



Figure 7-1: A340-541 in ascent: Variation of the MAE in fuel flow rate prediction with deviation in the estimated TOW, for unseen data not used for model training. The MAE is calculated using a GPR model.

Table 7.1 tabulates the percentage change in mean absolute error in fuel flow rate prediction in climb out and approach due to a +3% deviation in estimated TOW from its actual value. The model predictions are seen to be sensitive to even a 3% deviation from the true TOW, with MAE changing by as much as about 19% for the A320-214 in climb out. This sensitivity of the MAE in fuel flow rate prediction to the deviation in the estimated TOW motivates the need to accurately predict the TOW, in order to accurately predict the fuel flow rate.

Table 7.1: Increase in fuel flow rate MAE for a +3% deviation in TOW from its actual value. Fuel flow rate GPR models are used in this analysis. Table entries are obtained by averaging across all flights in the combined validation and test datasets (i.e., unseen data not used for training).

| A/C Type | Phase | % Increase in MAE |
|---|---|---|
| A319-112 | Climb out | 12.9 |
| | Approach | 5.9 |
| A320-214 | Climb out | 19.2 |
| | Approach | 5.2 |
| A321-111 | Climb out | 18.0 |
| | Approach | 4.9 |
| A330-202 | Climb out | 15.0 |
| | Approach | 4.3 |
| A330-243 | Climb out | 3.9 |
| | Approach | 3.1 |
| A340-541 | Climb out | 6.7 |
| | Approach | 1.0 |
| B767-300 | Climb out | 12.8 |
| | Approach | 9.7 |
| B777-300ER | Climb out | 4.7 |
| | Approach | 3.1 |

## 7.2 Model features

Figure 7-2 shows the free-body diagram of an aircraft during takeoff roll on the runway with the forces acting on it. The equations of motion during takeoff ground roll are given in Equations 7.1–7.7.

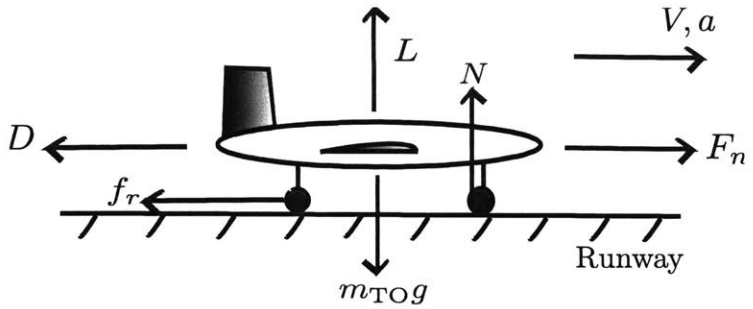



Figure 7-2: Schematic of airplane dynamics during takeoff ground roll.

$$L + N = m_{\mathrm{TO}}g \tag{7.1}$$

$$F_n - D - f_r = m_{\mathrm{TO}}a \tag{7.2}$$

$$L = q\mathcal{S}C_L \tag{7.3}$$

$$D = q\mathcal{S}C_D \tag{7.4}$$

$$f_r = \mu_r N \tag{7.5}$$

$$a = \frac{dV}{dt} \tag{7.6}$$

$$q = \frac{1}{2}\rho_\infty V^2 \tag{7.7}$$

Winds are neglected in this analysis. Neglecting wind speeds during takeoff ground roll, the aircraft airspeed is assumed to be equal to the ground speed ($V = V_{\mathrm{GS}}$). The mass of fuel consumed during the takeoff ground roll is assumed to be small compared to the aircraft mass, so that the aircraft weight is effectively constant and equal to the TOW ($m_{\mathrm{TO}}$) during the takeoff ground roll. The coefficients of lift and drag, governed by the aircraft configuration, are also assumed to be constant during this flight phase. The net thrust on the aircraft is the averaged net thrust per engine times the number of engines ($N_{\mathrm{eng}}$). The net thrust per engine is assumed to be a function of the static thrust ($F_0$) and the aircraft velocity [68]. The static thrust is the net thrust which would be produced by the engine if the aircraft were at rest at the set throttle setting. During the takeoff roll, the throttle setting does not change. The static thrust is assumed to be a function of the thrust deration level ($\eta$), the ambient air density during the takeoff roll ($\rho_\infty$) and the maximum sea level, static engine thrust ($F_{00}$). The net thrust on the aircraft is

therefore, governed by the following functional relation:

$$F_n = f_{F_n}(N_{\text{eng}}, V_{\text{GS}}, \eta, \rho_\infty, F_{00}) \tag{7.8}$$

The distance covered during takeoff roll ($S$) can be calculated by the following equation:

$$S = \int_{V_{\text{GS}_1}}^{V_{\text{GS}_2}} V_{\text{GS}} \frac{dV_{\text{GS}}}{a} \tag{7.9}$$

Here, $V_{\text{GS}_1}$ is the aircraft ground speed at the start of the takeoff ground roll and $V_{\text{GS}_2}$ is the aircraft ground speed at wheels-off at the end of the takeoff ground roll. Combining Equations (7.1)–(7.9), the TOW can be expressed by the following functional relation:

$$m_{\text{TO}} = f_{m_{\text{TO}}}(S, \rho_\infty, V_{\text{GS}_1}, V_{\text{GS}_2}, \mathcal{S}, F_{00}, C_L, C_D, \mu_r, \eta, N_{\text{eng}}) \tag{7.10}$$

$\mathcal{S}$, $F_{00}$, and $N_{\text{eng}}$ are constants for a given aircraft/engine type.

The modeling variables are now restricted to only those which can be obtained or derived from easily accessible databases. The ground roll distance and aircraft ground speed during ground roll can be derived from surface surveillance data, while the ambient air density can be obtained from airport weather data. By contrast, the values of the aircraft lift and drag coefficients, coefficient of friction, and thrust deration level are difficult to obtain, and are therefore, not used as model features. Hence, for a particular aircraft type, the model uses the ground roll distance ($S$, in m), the ambient air density ($\rho_\infty$, in kg·m$^{-3}$) during roll, the aircraft ground speed at the start of the takeoff ground roll ($V_{\text{GS}_1}$, in m·s$^{-1}$), and the aircraft ground speed at the end of the takeoff ground roll ($V_{\text{GS}_2}$, in m·s$^{-1}$) as the predictor/input variables. The predicted/output variable is the aircraft TOW ($m_{\text{TO}}$, in kg). In other words, the TOW prediction model has the following form:

$$m_{\text{TO}}\Big|_{\text{aircraft type}} \approx m_{\text{TO}}(S, \rho_\infty, V_{\text{GS}_1}, V_{\text{GS}_2}) \tag{7.11}$$

The unmodeled features will contribute to the uncertainty of the TOW estimate, and will be reflected in the prediction intervals provided by the statistical models.

## 7.3 Model training

Ground truth FDR values of the predictor and the predicted variables during takeoff roll in the training dataset are used for model training. Each observation (data point) in the training, validation and test sets for TOW modeling corresponds to the takeoff of one flight. Prior to training, all the variables are standardized, that is shifted by the sample mean and then scaled by the sample standard deviation of the respective variables in the training datasets. Due to its performance seen in fuel flow rate modeling, Gaussian Process Regression (GPR) is also used to train the TOW models. A validation study using the validation datasets is done to choose the appropriate kernel functions to build the final GPR models to predict the TOW. Once the models are trained, they can be used to determine the point estimates, the prediction intervals, and the predictive distributions of the TOW for a new input vector. Under GPR, the TOW predictive distribution is a normal/Gaussian distribution.

## 7.4 Model results and comparisons with other models

Table 7.2 shows the performance of the GPR models in predicting the TOW on the test datasets for the different aircraft types. The table also shows the performance of the TOW estimation model given by the Aircraft Noise and Performance (ANP) database. A part of AEDT, the ANP database models the TOW as a piecewise constant function of the flight stage/trip length [10]. The flight stage/trip length is determined in this chapter by calculating the great circle distance between the flight origin and destination airports. The ANP model is a deterministic model and does not account for operational variability in TOW. Two aircraft flying the same trip length may have different TOW due to operational reasons such as fuel tankering. Since the ANP model is a deterministic model, its PC and NLPI are not reported in the table. Table 7.2 also shows the statistical comparison of the model MEs with zero and the model MAEs with one another. For comparison of MEs, the null hypothesis is that the MEs are zero (unbiased model predictions). The alternate hypothesis is that the MEs are non-zero (biased model predictions). For comparison of MAEs, the null hypothesis is that the GPR model gives a similar (or worse) predictive performance as compared to the ANP model in terms of a similar (or higher) MAE.

The alternate hypothesis is that the GPR model gives a better predictive performance than the ANP model in terms of a lower MAE (i.e., $MAE_{GPR} < MAE_{ANP}$). The ME and the MAE used for these statistical comparison studies are calculated on a per-flight basis (and not across flights) for a particular aircraft type. Table 7.2 shows the p-values obtained through the Wilcoxon signed-rank test (Section 4.6). The p-values indicating acceptance of the alternate hypothesis at the 5% significance level are highlighted in bold.

Table 7.2: TOW estimation: Performance metrics of the GPR models, and the ANP models, on the test datasets for different aircraft types. All the evaluation metrics are calculated on de-standardized data. The table also shows the p-values for statistical comparisons of different model metrics (i.e., $ME_{GPR} \neq 0$, $ME_{ANP} \neq 0$, and $MAE_{GPR} < MAE_{ANP}$). Each cell (except that for a p-value) shows the mean value (and the standard deviation) of the particular metric. All entries (except the p-values) are in percent. The data for the A330-343 are drawn from the FDR-II dataset. The data for all the other aircraft types are drawn from the FDR-I dataset.

| A/C Type | ME | | MAE | | PC | NLPI | Model Comparison: p-values | | |
|---|---|---|---|---|---|---|---|---|---|
| | GPR | ANP | GPR | ANP | GPR | GPR | $ME_{GPR}$ | $ME_{ANP}$ | MAE |
| A319-112 | 0.6 (5.9) | 1.0 (6.3) | 5.0 (3.0) | 5.3 (3.3) | 84.6 (36.8) | 18.4 (0.9) | 0.657 | 0.439 | 0.319 |
| A320-214 | 0.4 (4.9) | 0.1 (5.1) | 3.9 (2.9) | 4.3 (2.7) | 97.1 (17.1) | 19.3 (0.9) | 0.778 | 0.871 | 0.199 |
| A321-111 | -0.3 (7.5) | 1.0 (8.5) | 6.1 (4.2) | 6.7 (5.1) | 91.3 (28.8) | 23.9 (1.0) | 0.808 | 1.000 | 0.303 |
| A330-202 | -1.8 (2.6) | -4.1 (5.6) | 2.4 (1.9) | 6.0 (3.3) | 94.1 (24.3) | 11.2 (2.1) | **0.025** | **0.013** | **0.007** |
| A330-243 | -0.4 (3.4) | 0.4 (6.5) | 2.6 (2.1) | 4.5 (4.5) | 85.0 (36.6) | 10.2 (1.3) | 0.681 | 0.737 | 0.058 |
| A330-343 | 0.2 (4.3) | 1.8 (4.1) | 3.6 (2.3) | 3.5 (2.7) | 97.3 (16.4) | 17.5 (0.8) | 0.874 | **0.021** | 0.666 |
| A340-541 | 2.5 (3.8) | 12.5 (14.0) | 4.0 (1.9) | 12.7 (14.0) | 90.0 (31.6) | 14.4 (9.4) | 0.093 | **0.009** | **0.023** |
| B767-300 | 0.2 (2.5) | -6.5 (6.7) | 2.0 (1.5) | 8.4 (3.8) | 94.4 (23.6) | 10.9 (1.8) | 0.528 | **0.004** | **1.4e-4** |
| B777-300 ER | 0.1 (3.4) | 3.5 (6.3) | 2.0 (2.7) | 5.6 (4.5) | 92.3 (27.2) | 8.2 (2.4) | 0.517 | **0.007** | **0.002** |

Figures 7-3 and 7-4 contain box-plots showing the errors and the absolute errors, respectively, for the TOW prediction across the different flights in test data. Model predictions are unbiased (at a 5% significance level) if the error box-plots in Figure 7-3 contain zero in their notched regions. The notched regions of the absolute error box-plots in Figure 7-4 for the GPR

models can be compared to those of the box-plots for the corresponding ANP models to determine which model gives a statistically significantly lower MAE (at a 5% significance level) on the TOW prediction.

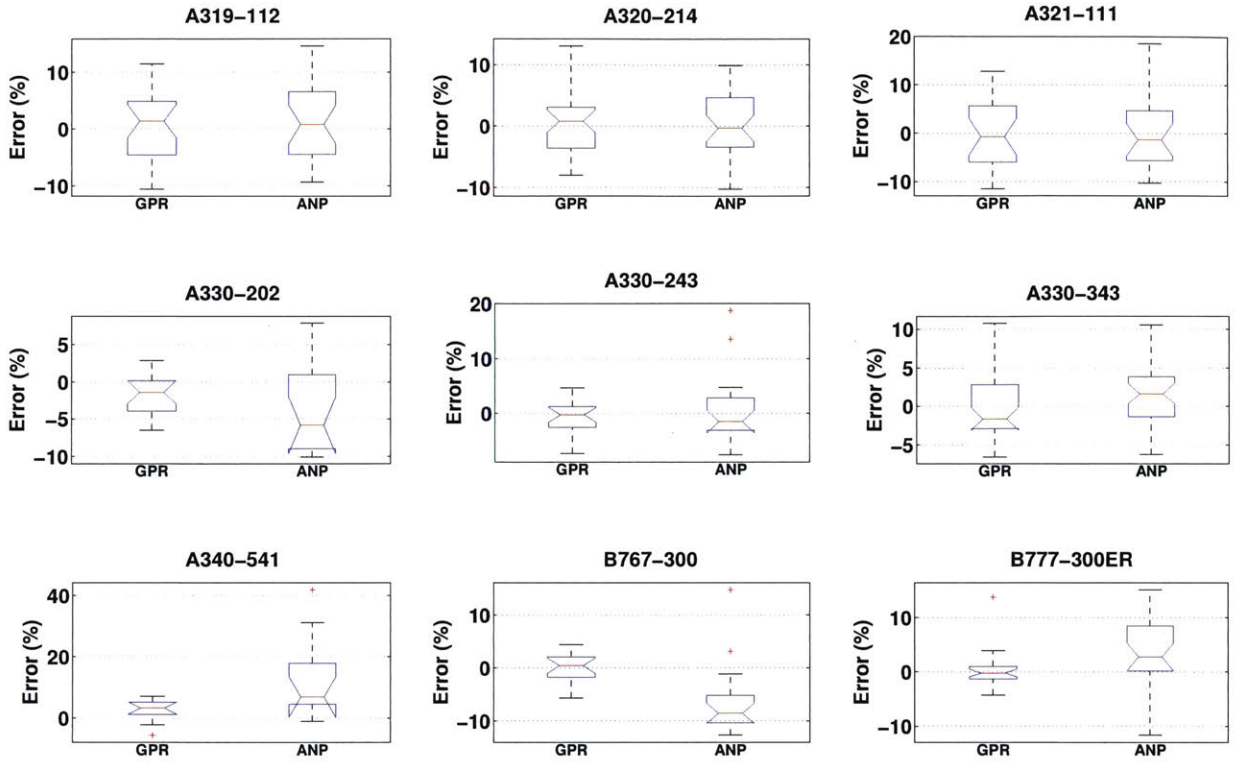



Figure 7-3: Box-plots showing error on TOW prediction for flights in test data. These box-plots are helpful in checking for model bias in the mean predictions.

From Table 7.2, it can be seen that the proposed GPR models give unbiased predictions of the TOW in most of the cases. The median of the MEs across the different aircraft types is 0.2%. The proposed GPR models give a median MAE of 3.6% across the different aircraft types whereas the ANP models give a median MAE of 5.6%. The median PC given by the GPR models is 92.3% across the different aircraft types. The median NLPI is 14.4%. The statistical comparison tests for the MAE indicate that the GPR models perform statistically significantly (at a 5% significance level) similar to or better than the ANP models for all the aircraft types in the study. The GPR models reduce the median MAE by 36% as compared to the ANP models. Moreover, unlike the ANP models, the GPR models also give prediction intervals which can quantify the uncertainty in TOW.

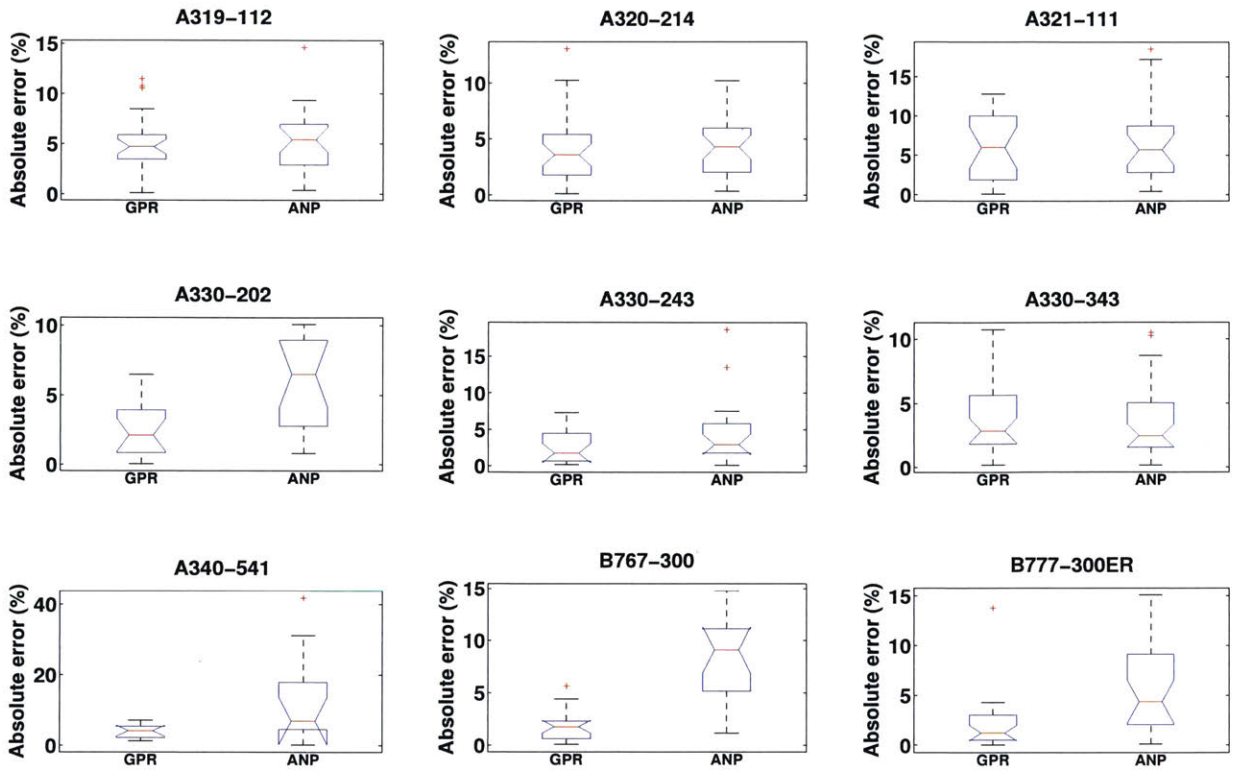



Figure 7-4: Box-plots showing absolute error on TOW prediction for flights in test data. These box-plots can be used to compare the GPR models with the corresponding ANP models for the accuracy of mean predictions.

## 7.5 ASDE-X data as inputs

In this section, TOW model predictive performance for the A330-343 is evaluated using values of the trajectory variables drawn from the ASDE-X dataset. The values of the evaluation metrics determined in this section indicate the predictive performance of the models as they will be used in practice using ground surveillance data.

### 7.5.1 Predictive performance with ASDE-X data inputs

Models trained on A330-343 FDR data in Section 7.3 are applied to trajectory predictor variable (i.e., ground speed and takeoff roll distance) values drawn from the ASDE-X data to determine mean predictions and prediction intervals for the TOW. Values of the ambient air density, and the ground truth TOW are still obtained from the corresponding FDR records for the A330-343.

Model evaluation is done on 33 flights for the A330-343 which have records in both the FDR-II and the ASDE-X datasets and which have not been used for TOW model training.

Table 7.3 tabulates the TOW predictive performance when the source of trajectory variables is ASDE-X. For reference, the predictive performance when the trajectory variables are obtained from FDR data is also tabulated. It should be noted that this predictive performance on FDR-obtained variables may differ from that recorded in Table 7.2 as different sets of flights have been used for model evaluation in the two tables.

Table 7.3: Performance of the GPR model and the ANP model for the A330-343 to predict TOW when the trajectory predictor variables are obtained from ASDE-X data. Each cell reports the mean and standard deviation (within parentheses) of the evaluation metric across all the flights in unseen data not used for model training. For reference, model predictive performance when the trajectory predictor variables are obtained from FDR data is also given.

| **Trajectory Source** | **ME (%)** | | **MAE (%)** | | **PC (%)** | **NLPI (%)** |
|---|---|---|---|---|---|---|
| | **GPR** | **ANP** | **GPR** | **ANP** | **GPR** | **GPR** |
| ASDE-X | 1.9 (4.1) | 1.3 | 3.6 (2.7) | 3.4 | 97.0 (17.4) | 17.5 (1.0) |
| FDR | 1.2 (3.8) | (4.0) | 3.3 (2.2) | (2.3) | 100.0 (0.0) | 17.4 (0.8) |

Table 7.3 shows that for the A330-343, model predictive performance (in terms of MAE, PC, NLPI) when trajectory variables are obtained from ASDE-X is similar to that when trajectory variables are obtained from FDR. The GPR model gives a statistically similar MAE as the ANP model. It should be noted that the ANP models do not need any trajectory variables as inputs, and are therefore independent of the source of such data.

## 7.6 Use of predicted takeoff weight to model fuel flow rate

In this section, TOW models developed in Section 7.3 are used to predict TOW which is further used as an input variable to predict the fuel flow rate. The aim of the analysis done in this section is to understand how the fuel flow rate model predictive performance is affected when an uncertain estimate of the TOW is used as an input instead of using the ground truth value of the TOW (as has been done in Chapter 5).

### 7.6.1 Methodology

Using the TOW predicted by the ANP models as well as by the developed GPR models (Section 7.3) as inputs to the fuel flow rate GPR models (Section 5.3.1), their predictive performance on the flights in the unseen prediction dataset (combined validation and test datasets) is now evaluated. To incorporate uncertainty in the predicted TOW, the fuel flow rate predictive performance is evaluated using the predictive distribution of the fuel flow rate marginalized over the uncertain values of TOW. In other words, we are interested in computing the following:

$$p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, \phi, \mathcal{D}_1, \mathcal{D}_2) = \int_{m_{\mathrm{TO}}} p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}}, \mathcal{D}_1) p(m_{\mathrm{TO}}|\phi, \mathcal{D}_2) dm_{\mathrm{TO}} \tag{7.12}$$

Here, $p$ refers to the Probability Distribution Function (PDF), $\dot{m}_f$ is the fuel flow rate to be predicted, $\mathbf{x}_{-\mathrm{TOW}}$ is the vector of predictor variables in the fuel flow rate GPR model excluding the TOW, $m_{\mathrm{TO}}$ is the TOW, and $\mathcal{D}_1$ is the set of the training variables used for building the fuel flow rate GPR model. $p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}}, \mathcal{D}_1)$ is the PDF of the predictive distribution given by the fuel flow rate GPR model when a known TOW is used as an input variable. It is thus, a Gaussian PDF. $p(m_{\mathrm{TO}}|\phi, \mathcal{D}_2)$ is the distribution of the predicted TOW parametrized by $\phi$ and $\mathcal{D}_2$.

The ANP model is a deterministic model giving a flight stage length-based point estimate of the TOW, $m_{\mathrm{TO,ANP}}$. Under the ANP model, Equation 7.12 becomes

$$p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, \phi, \mathcal{D}_1, \mathcal{D}_2) = p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, m_{\mathrm{TO,ANP}}, \mathcal{D}_1) \tag{7.13}$$

which is the PDF of a normal distribution under the GPR formulation.

The GPR models for TOW prediction (Section 7.3) give the complete predictive distribution for the TOW (which is a normal distribution). Therefore, under these GPR models for TOW, Equation 7.12 becomes

$$p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, \phi, \mathcal{D}_1, \mathcal{D}_2) = \int_{m_{\mathrm{TO}}} p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}}, \mathcal{D}_1) p(m_{\mathrm{TO}}|\phi, \mathcal{D}_2) dm_{\mathrm{TO}} \tag{7.14}$$

$$\approx \frac{1}{n_s} \sum_{i=1}^{n_s} p(\dot{m}_f|\mathbf{x}_{-\mathrm{TOW}}, m_{\mathrm{TO}_i}, \mathcal{D}_1). \tag{7.15}$$

When GPR models are used to predict the TOW, $\phi$ and $\mathcal{D}_2$ hold specific meanings. $\phi$ is the vector of predictor variables used in the GPR TOW prediction models (i.e., $S, \rho_\infty, V_{\mathrm{GS}_1}$, and $V_{\mathrm{GS}_2}$ as mentioned in Section 7.2). $\mathcal{D}_2$ is the training set used to build the GPR models to predict the TOW. Equation 7.15 approximates Equation 7.14 through a Monte Carlo approximation with $n_s$ samples of the TOW drawn from its Gaussian predictive distribution given by the GPR models for TOW prediction. In this study, $n_s$ is chosen to be 1,000. Equation 7.15 therefore, shows that the desired predictive distribution of the fuel flow rate under a GPR model of the TOW can be approximately modeled as a Gaussian Mixture distribution with $n_s$ equally weighted components.

### 7.6.2 Results

Depending on how the TOW variable is predicted, there are two variants of the GPR models developed to predict the fuel flow rate:

1. **Model 1:** This variant predicts the TOW predictor variable, that is input to the GPR fuel flow rate models, using the ANP method; and

2. **Model 2:** This variant predicts the TOW predictor variable using the GPR models developed in Section 7.3 for TOW prediction.

The predictive distributions for the fuel flow rate marginalized over the TOW for Model 1 (Equation 7.13), and Model 2 (Equation 7.15) are used to calculate the mean predictions and the 95% highest density prediction intervals for the fuel flow rates. Table 7.4 tabulates the predictive performance of these models on the unseen prediction dataset (combined validation and test datasets) in climb out and approach for different aircraft types. Statistical multiple

comparison tests (at a significance level of 5%) have been performed to determine which of the two models performs better on different evaluation metrics.

Table 7.4: Fuel flow rate model predictive performance using estimated TOW. MAE, PC, NLPI for the fuel flow rate predictions are determined in climb out and approach using unseen prediction data (combined validation and test data). Each cell entry records the mean and the standard deviation (within parentheses) of the evaluation metric.

| A/C Type | Phase | MAE (%) | | PC (%) | | NLPI (%) | |
|---|---|---|---|---|---|---|---|
| | | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 |
| A319-112 | Climb out | 3.8 (1.5) | 3.9 (1.8) | 83.9 (16.3) | 92.4 (12.1) | 13.6 (0.5) | 17.0 (1.2) |
| | Approach | 16.8 (5.4) | 16.8 (5.4) | 95.0 (4.8) | 95.7 (4.3) | 94.3 (12.1) | 96.5 (12.5) |
| A320-214 | Climb out | 4.2 (2.2) | 4.3 (2.4) | 91.6 (14.5) | 96.0 (9.0) | 20.9 (1.3) | 25.1 (2.2) |
| | Approach | 16.2 (6.3) | 16.6 (6.2) | 94.6 (4.9) | 95.0 (4.2) | 104.6 (95.2) | 99.3 (51.8) |
| A321-111 | Climb out | 7.8 (4.4) | 6.5 (2.9) | 72.5 (25.0) | 93.7 (10.0) | 21.2 (1.2) | 29.8 (1.7) |
| | Approach | 18.1 (4.7) | 17.2 (4.5) | 91.5 (5.5) | 93.4 (5.0) | 76.1 (11.0) | 80.8 (12.6) |
| A330-202 | Climb out | 5.1 (1.9) | 3.6 (1.5) | 83.1 (11.9) | 92.4 (8.2) | 19.7 (1.4) | 19.9 (1.4) |
| | Approach | 27.3 (10.1) | 28.0 (11.3) | 91.0 (7.1) | 91.1 (7.3) | 146.9 (22.1) | 141.7 (24.4) |
| A330-243 | Climb out | 3.0 (1.9) | 2.8 (2.0) | 89.8 (19.6) | 90.8 (20.7) | 12.6 (0.7) | 12.9 (0.8) |
| | Approach | 20.8 (11.5) | 20.6 (11.5) | 90.8 (9.0) | 91.1 (9.0) | 111.3 (20.6) | 110.8 (21.3) |
| A340-541 | Climb out | 9.1 (12.3) | 5.1 (5.4) | 66.6 (32.9) | 83.9 (22.4) | 13.9 (0.5) | 17.1 (4.9) |
| | Approach | 19.6 (5.5) | 17.6 (3.3) | 95.7 (3.9) | 96.5 (2.7) | 120.9 (22.5) | 120.8 (18.1) |
| B767-300 | Climb out | 5.3 (2.4) | 3.0 (1.5) | 83.2 (20.0) | 98.1 (4.3) | 19.2 (1.2) | 19.5 (2.0) |
| | Approach | 20.0 (7.4) | 19.8 (6.7) | 93.3 (7.5) | 95.6 (5.2) | 147.8 (58.6) | 147.0 (212.6) |
| B777-300ER | Climb out | 8.1 (4.3) | 6.7 (2.5) | 87.0 (10.1) | 93.0 (7.7) | 29.0 (3.4) | 31.8 (5.8) |
| | Approach | 17.4 (5.6) | 16.5 (5.3) | 93.8 (5.6) | 94.1 (5.2) | 99.5 (17.0) | 101.8 (17.9) |

In climb out, the ANP-TOW prediction model (Model 1) gives a fuel flow rate median MAE, PC, and NLPI of 5.2%, 83.6%, and 19.5%, respectively. The GPR-TOW prediction model (Model 2) gives a fuel flow rate median MAE, PC, and NLPI of 4.1%, 92.7%, and 19.7%, respectively. For the majority of the aircraft types, Model 2 gives a lower or similar MAE as compared to Model 1, and a higher or similar PC as compared to Model 1. Thus, for the majority of the aircraft types, using the GPR-TOW model gives a better or similar fuel flow rate predictive performance as compared to the ANP-TOW prediction model. However, the NLPI for Model 2 is greater than or equal to that for Model 1. This is expected as Model 2 propagates uncertainty in TOW to that in fuel flow rate and hence, leads to bigger prediction intervals.

In approach, the ANP-TOW prediction model (Model 1) gives a fuel flow rate median MAE, PC, and NLPI of 18.9%, 93.6%, and 108.0%, respectively. The GPR-TOW prediction model (Model 2) gives a fuel flow rate median MAE, PC, and NLPI of 17.4%, 94.6%, and 106.3%, respectively. For the majority of the aircraft types, Model 2 gives a lower or similar MAE as compared to Model 1, and a higher or similar PC as compared to Model 1. Thus, for the majority of the aircraft types, using the GPR-TOW model gives a better or similar fuel flow rate predictive performance as compared to the ANP-TOW prediction model. However, the NLPI for Model 2 is greater than or equal to that for Model 1 for the majority of the aircraft types. This is again expected as Model 2 propagates uncertainty in TOW to that in fuel flow rate and hence, leads to bigger prediction intervals.

It should be noted that the difference in fuel flow rate predictive performance for the GPR-TOW and the ANP-TOW prediction models in climb out is starker than that observed in approach. There is a greater number of aircraft types giving similar performance from the two models in approach as compared to climb out. This observation can also be made from Table 7.1 where the increase in fuel flow rate MAE is greater in climb out than in approach for a similar deviation of the estimated TOW from its true value.

Finally, Table 7.5 compares the fuel flow rate predictive performance in climb out and approach when exact TOW values (Section 5.3.1) are used as inputs to that when GPR-estimated TOW values (Equation 7.15) are used as inputs. In the majority of the cases, using exact TOW

Table 7.5: Comparison of fuel flow rate model predictive performance using exact and estimated TOW. MAE, PC, NLPI for the fuel flow rate predictions are determined in climb out and approach using unseen prediction data (combined validation and test data). Each cell entry records the mean and the standard deviation (within parentheses) of the evaluation metric.

| A/C Type | Phase | MAE (%) | | PC (%) | | NLPI (%) | |
|---|---|---|---|---|---|---|---|
| | | Exact TOW | GPR-Estimated TOW | Exact TOW | GPR-Estimated TOW | Exact TOW | GPR-Estimated TOW |
| A319-112 | Climb out | 2.8 (1.3) | 3.9 (1.8) | 92.7 (11.8) | 92.4 (12.1) | 14.0 (1.0) | 17.0 (1.2) |
| | Approach | 16.7 (5.9) | 16.8 (5.4) | 95.3 (4.7) | 95.7 (4.3) | 95.9 (13.6) | 96.5 (12.5) |
| A320-214 | Climb out | 3.5 (2.1) | 4.3 (2.4) | 95.1 (9.0) | 96.0 (9.0) | 21.0 (1.3) | 25.1 (2.2) |
| | Approach | 16.4 (6.3) | 16.6 (6.2) | 95.1 (4.0) | 95.0 (4.2) | 71.4 (155.5) | 99.3 (51.8) |
| A321-111 | Climb out | 4.3 (1.4) | 6.5 (2.9) | 92.2 (8.0) | 93.7 (10.0) | 21.7 (1.1) | 29.8 (1.7) |
| | Approach | 16.8 (4.4) | 17.2 (4.5) | 92.8 (5.6) | 93.4 (5.0) | 79.0 (14.8) | 80.8 (12.6) |
| A330-202 | Climb out | 4.1 (1.6) | 3.6 (1.5) | 92.0 (9.5) | 92.4 (8.2) | 18.2 (1.1) | 19.9 (1.4) |
| | Approach | 27.3 (10.9) | 28.0 (11.3) | 91.3 (7.5) | 91.1 (7.3) | 145.9 (26.7) | 141.7 (24.4) |
| A330-243 | Climb out | 3.1 (1.9) | 2.8 (2.0) | 91.6 (17.4) | 90.8 (20.7) | 12.6 (0.9) | 12.9 (0.8) |
| | Approach | 21.2 (11.8) | 20.6 (11.5) | 91.0 (9.3) | 91.1 (9.0) | 109.2 (20.1) | 110.8 (21.3) |
| A340-541 | Climb out | 4.1 (2.7) | 5.1 (5.4) | 83.7 (22.7) | 83.9 (22.4) | 14.4 (0.9) | 17.1 (4.9) |
| | Approach | 17.6 (3.4) | 17.6 (3.3) | 95.4 (3.7) | 96.5 (2.7) | 123.8 (20.0) | 120.8 (18.1) |
| B767-300 | Climb out | 2.9 (1.5) | 3.0 (1.5) | 97.5 (4.5) | 98.1 (4.3) | 17.7 (1.4) | 19.5 (2.0) |
| | Approach | 19.0 (6.2) | 19.8 (6.7) | 95.5 (5.2) | 95.6 (5.2) | 121.6 (46.5) | 147.0 (212.6) |
| B777-300ER | Climb out | 6.5 (1.7) | 6.7 (2.5) | 91.7 (9.2) | 93.0 (7.7) | 31.6 (5.8) | 31.8 (5.8) |
| | Approach | 17.2 (4.9) | 16.5 (5.3) | 94.4 (4.7) | 94.1 (5.2) | 99.8 (16.3) | 101.8 (17.9) |

values gives a statistically significantly similar or better fuel flow rate predictive performance. In climb out, using estimates of TOW as inputs leads to an up to 51.2% increase in MAE, and up to 37.3% increase in NLPI as compared to using exact values of TOW. Similarly in approach, using estimates of TOW as inputs leads to an up to 2.3% increase in NLPI as compared to using exact values of TOW. Thus, the practical implementability of models which first estimate TOW and then use the TOW estimate as an input to further estimate the fuel flow rate is expectedly accompanied with a drop in the fuel flow rate predictive performance.

# Chapter 8

# Conclusions

This thesis developed a data-driven methodology to model aircraft engine fuel flow rate and fuel burn using a knowledge of its trajectory. The methodology resulted in models which can give mean predictions as well as prediction intervals for the fuel flow rate and the mass of fuel consumed. This chapter summarizes the key findings of the thesis, and describes some promising directions for future research.

## 8.1 Summary

The main objective of this thesis was to develop a framework to model aircraft engine fuel flow rate and fuel burn for a flight, given its trajectory. The fuel flow rate was directly mapped to the aircraft trajectory using statistical data-driven models, thereby bypassing many intermediate parameters such as drag, net thrust. Operational data from flight data recorders were used for model identification, enabling the development of models that were representative of real flight operations. The thesis focused on the problem of predicting both the instantaneous fuel flow rates, as well as the total mass of fuel consumed in different phases of flight. Statistical algorithms were applied to operational data, and enabled the generation of mean estimates as well as prediction intervals for the fuel flow rates. These prediction intervals quantified the cumulative extent of operational variability in fuel fuel rates as well as model uncertainty. Different regression approaches were investigated, and Gaussian Process Regression (GPR) was found to

give good predictive performance. An understanding of the aircraft dynamics was leveraged to select features relevant for model development. To enable practical implementation, only those features which were derivable from trajectory data were retained as input variables.

### 8.1.1 Airborne fuel flow rate modeling

Different models were built to model the fuel flow rate in the different airborne phases (viz., ascent, climb out, cruise, descent, approach) for each of 8 aircraft types. Thus, the models are local experts for each phase and are expected to give better predictive performance than a single model built for the entire flight. Validation studies using different statistical algorithms revealed the GPR-based batch prediction algorithm to give the best fuel flow rate predictive performance. Input features for these models included the aircraft dynamic pressure multiplied by the reference wing area, the aircraft takeoff weight, the ratio of the aircraft vertical speed to its ground speed, the ground speed, the rate of change of ground speed, and the aircraft altitude above the arrival airport elevation. These GPR models gave the entire predictive distribution of the fuel flow rate given an input vector. Being nonparametric, GPR does not require a choice of the form of the regression features prior to model training. The regression functions can therefore, adapt themselves during model training to the system complexity represented by performance data. The GPR-based batch prediction algorithm was evaluated using different evaluation metrics on unseen test data not used for model training. The metrics evaluated the statistical models for the accuracy of their mean and interval predictions. The GPR-based batch prediction models (using a known TOW) gave a median value of the mean error (ME) in the fuel flow rates of 0.8% in ascent, 0.4% in climb out, 2% in cruise, 7.5% in descent, and 6.4% in approach across the different aircraft types on test data. The median value of the mean absolute error (MAE) across the different aircraft types was 4.6% in ascent, 3.8% in climb out, 10.9% in cruise, 22.4% in descent, and 18.0% in approach. The GPR models were compared for their predictive performance with current state-of-the-practice APMs such as Base of Aircraft Data (BADA), the Senzig-Fleming-Iovinelli models, and the ICAO Databank models with Boeing Fuel Flow Method 2 corrections - all components of the FAA's AEDT used for fuel burn, emissions, and noise modeling. Unlike most other studies which tend to compare models on the basis of the

mean values of the evaluation metrics, in this thesis, models were statistically compared so that the variance in the metrics across different flights is also taken into account. These statistical comparisons showed that the GPR models gave a lower MAE on the fuel flow rates as compared to the other APMs. The GPR models achieved a reduction in median MAE of as much as 48% in ascent, 71% in climb out, 49% in cruise, 31% in descent, and 77% in approach. Moreover, currently used APMs focus only on point estimation of fuel flow rates, and do not quantify the uncertainty therein. The GPR models gave the complete predictive distribution of the fuel flow rates which can be used to generate 95% prediction intervals. These prediction intervals gave a median value of the prediction coverage (PC) of 95% in ascent, 91.8% in climb out, 96.1 % in cruise, 93.2% in descent, and 94.4% in approach across the different aircraft types on test data. The closeness of these empirically observed PCs to 95% indicated good model specification. The simultaneous emphasis on mean estimation as well as uncertainty estimation is a key highlight of this thesis.

The Gaussian predictive distributions of the fuel flow rates using the GPR models led to the fuel burn (i.e., the mass of fuel consumed in a phase of flight) being modeled through Gaussian Mixture distributions. The GPR-based models gave a median ME in the fuel burn across the different aircraft types (using a known TOW) of 0.4% in ascent, 0.3% in climb out, -0.1% in cruise, 3.7% in descent, and 2.0% in approach. The median MAE in fuel burn given by these GPR models was 2.2% in ascent, 2.0% in climb out, 7.2% in cruise, 8.2% in descent, and 5.5% in approach. The GPR models achieved a reduction in median MAE of as much as 54% in ascent, 80% in climb out, 34% in cruise, 66% in descent, and 89% in approach as compared to currently used APMs. The proposed GPR-based models achieved a median value of the mean error of 0.2% (across flights of 8 different aircraft types), and a median value of the mean absolute error of 3.8% in the total airborne fuel burn. These results constitute a significant improvement over existing fuel burn models (for example, a nearly 20% improvement over the BADA model).

### 8.1.2 Fuel flow rate during taxi-out

On the airport surface, the fuel flow rate was modeled as a function of the ambient temperature, ambient pressure, taxi ground speed, takeoff mass, and the taxi acceleration. Examination of fuel flow rate profiles led to the identification of two distinct regions: A baseline fuel flow region and a fuel flow spike region. Since a very large part of the taxi-out fuel consumption was observed to occur during the baseline fuel flow region, only the baseline fuel flow was considered for modeling in this thesis. Two different models were developed to predict the baseline fuel flow rate: An OLS regression -based parametric model for the mean baseline fuel flow rate, and a GPR-based nonparametric model for the instantaneous baseline fuel flow rate. Both the models were found to give similar ME (about -3% for the A330-343 and -1.8% for the B777-300ER) and MAE (6.3% for the A330-343 and 2.7% for the B777-300ER) in taxi fuel flow rate (using a known TOW). Thus, both the models gave similar mean predictions for the fuel flow rate. When compared to the AEDT models, the statistical models developed were found to give a reduction in MAE of 84% for the A330-343 and 94% for the B777-300ER. The GPR-based models gave a higher prediction coverage due to their ability to model variability in the baseline fuel flow as compared to the OLS regression-based models which modeled only the mean of the baseline fuel flow rates. The GPR-based models could thus capture uncertainty in baseline fuel flow more accurately as compared to the OLS regression-based models. In terms of the taxi fuel burn predictions, the statistical models gave an ME of -4.7% (MAE 4.9%) for the A330-343 and an ME of -3.0% (MAE 3%) for the B777-300ER, achieving a reduction in MAE of as much as 93% as compared to the AEDT models. The practical utility of the statistical models developed was successfully demonstrated by evaluating model predictive performance using trajectory predictor variables from ground-based surveillance data sources (such as ASDE-X).

### 8.1.3 Takeoff weight predictions

Takeoff Weight (TOW) was identified as a predictor variable for modeling fuel flow. Fuel flow rate predictions were shown to be sensitive to TOW. Since the operational value of TOW on a per-flight basis is not available, statistical models were developed to estimate TOW using

trajectory information during takeoff ground roll. GPR-based models using ground speeds at the start and the end of takeoff ground roll, ambient density, and ground roll distance as input variables were found to give a TOW predictive performance similar to or better than that given by the ANP models used in AEDT for TOW estimation. The GPR-based models gave a median TOW prediction ME of 0.2% and MAE of 3.6%. The statistical models were able to account for uncertainty in TOW prediction and gave a median prediction coverage of 92.3%. When model TOW estimates were used as inputs to the fuel flow rate models, it was observed that the GPR-based models of TOW prediction gave a fuel flow rate predictive performance similar to or better than the ANP models for the majority of the aircraft types studied. In climb out, the median MAE and PC across the different aircraft types (using TOW estimated by GPR) were 4.1% and 92.7% respectively. In approach, the median MAE and PC across the different aircraft types (using TOW estimated by GPR) were 17.4% and 94.6% respectively. Using the TOW prediction models along with the fuel flow rate models developed in this thesis results in a framework to predict the fuel flow rate solely from trajectory information.

## 8.2 Opportunities for future research

The results presented in this thesis indicate the significant potential of data-driven models of aircraft engine performance. They also suggest interesting future research directions, both in terms of model refinement, and the application of similar approaches to other related problems.

### 8.2.1 Model refinements and extensions

During the course of this research, we identified potential model refinements that could be investigated in future research.

#### Models for a wider range of aircraft operations

This study was restricted to the modeling of nine aircraft types, considering operations at only 89 airports. In addition, these airports were of limited geographical diversity, being primarily located in Europe. As in the case of any data-driven approach, the developed models may not

be valid if the inputs vary significantly from those used in the training dataset. As a result, the models are likely to be representative of a small fraction of global aircraft operations. An important direction of future research would be to develop models that span a wider range of aircraft and engine types, as well as operating environments. The framework and methodologies proposed in this thesis can be easily extended to do so.

### Modeling airline-specific variations

The data used in this study were obtained primarily from one airline (although they included flights from two airlines, geographically dispersed). In practice, some variability is to be expected due to differences in airline operating procedures. An interesting topic of research would be to build airline-specific fuel burn models, and to analyze the differences between them.

### Developing generalized models

Each model developed in this thesis is specific to a particular aircraft and engine type combination. Future work can look into developing generalized models which can estimate the fuel flow and fuel burn for a family of aircraft/engine types. This will reduce the total number of models needed to be developed. As discussed in Section 5.7, one way to develop generalized models would be to train them using a combined dataset from a family of similarly performing aircraft variants (such as the A320 family, or the A330-200 family).

### Incorporating high-fidelity weather data

The models considered in this thesis used the ground speed as an input variable, and did not explicitly consider the impact of winds. Additionally, the ambient temperature and pressure were assumed to vary according to the ISA model. In practice, winds (especially the jet stream during cruise) are likely to affect the trajectory and fuel burn of individual flights. Sensitivity studies using AEDT indicate that headwind values have a large contribution to variability in fuel consumption and emissions [111]. As noted in Chapter 5, current archives of weather are not of the same spatial and temporal resolution as trajectory data, and were therefore, not considered as features in the models developed to date. A potential topic of future research involves

evaluating the benefits that could be achieved through the incorporation of approximate weather observations as inputs. Such inputs could be obtained, for example, by interpolating 3-hourly weather predictions given by the Global Forecast System of the National Centers for Environmental Prediction [112]. Over the next few years, it is expected that aircraft will broadcast wind observations, resulting in higher-fidelity wind data.

### Prediction of total fuel burn

This thesis focused on determining statistical models of fuel flow rate at each instant along the trajectory. The fuel flow rates were then aggregated to predict the total fuel burn of the flight. As mentioned in Chapter 5, if the only desired output is the total fuel burn, it may be possible to improve prediction performance by treating the total fuel burn as the predicted output (rather than the fuel flow rate). It is expected that more precise estimates of fuel burn can be obtained by directly modeling the fuel burn as the output variable of interest. This problem would be an interesting question for future research.

### Additional features and model refinements

The models can also be refined through the incorporation of additional features. For example, the great circle distance could be included as a feature for TOW prediction. Another improvement would be the automated identification of flight phases. In this thesis, the flight trajectory was divided into the standard flight phases (taxi, takeoff roll, ascent, cruise, descent) through a manual identification (that is, with pre-specified conditions) of change points. The effect of aircraft and engine degradation on fuel burn can be modeled by including the age of the aircraft or the engine, if available, as a predictor variable. Finally, more accurate estimates of fuel flow rates and fuel burn in the taxi-out phase can be obtained by modeling the fuel flow spike regions, in addition to the baseline fuel flow rates. BADA Family 3 was considered in the comparative analyses conducted in this thesis, because of the wide range of aircraft types it includes. Future work would study comparisons with the BADA Family 4 models as well.

### 8.2.2 Extensions to other aircraft and engine performance characteristics

Finally, this thesis focused on fuel burn as a quantity of interest. However, the methods presented in this thesis can be easily extended to other aspects of aircraft performance, such as thrust, aircraft landing weight, etc. These would be important topics for future research.

# Appendix A

# FDR Data: Fields

Table A.1: FDR-I data fields.

| Sr. No. | Description | Unit in FDR | SI Unit |
|---|---|---|---|
| 1 | Flight record (given) | - | - |
| 2 | Fleet (given) | - | - |
| 3 | Average $N_1$ over all engines from start of event (% of maximum) | - | - |
| 4 | Average $N_1$ left inboard engine from start of event (% of maximum) | - | - |
| 5 | Average $N_1$ left outboard engine from start of event (% of maximum) | - | - |
| 6 | Average $N_1$ right inboard engine from start of event (% of maximum) | - | - |
| 7 | Average $N_1$ right outboard engine from start of event (% of maximum) | - | - |
| 8 | Average burner pressure $P_3$ over all engines from start of event | psi | Pa |
| 9 | Average $P_3$ left inboard engine from start of event | psi | Pa |
| 10 | Average $P_3$ left outboard engine from start of event | psi | Pa |
| 11 | Average $P_3$ right inboard engine from start of event | psi | Pa |
| 12 | Average $P_3$ right outboard engine from start of event | psi | Pa |
| 13 | Average total fuel flow all engines from start of event | kg/h | kg/s |
| 14 | Average fuel flow left inboard engine from start of event | kg/h | kg/s |
| 15 | Average fuel flow left outboard engine from start of event | kg/h | kg/s |
| 16 | Average fuel flow right inboard engine from start of event | kg/h | kg/s |
| 17 | Average fuel flow right outboard engine from start of event | kg/h | kg/s |
| 18 | Mean true airspeed (TAS) from start of event (sample interval) | knot | m/s |
| 19 | Mean ground speed (GS) from start of event (sample interval) | knot | m/s |

| Sr. No. | Description | Unit in FDR | SI Unit |
|---|---|---|---|
| 20 | Mean vertical speed (inertial) from start of event (sample interval) | ft/min | m/s |
| 21 | Mean Mach number from start of event (sample interval) | - | - |
| 22 | Average atmospheric pressure (ambient, undisturbed air, sample interval) | hPa | Pa |
| 23 | Dynamic pressure (ambient, undisturbed air, sample interval) | hPa | Pa |
| 24 | Mean lateral acceleration (sample interval) | g | $m/s^2$ |
| 25 | Maximum lateral acceleration (sample interval) | g | $m/s^2$ |
| 26 | Mean longitudinal acceleration (sample interval) | g | $m/s^2$ |
| 27 | Maximum longitudinal acceleration (sample interval) | g | $m/s^2$ |
| 28 | Mean normal load factor (sample interval) | g | - |
| 29 | Maximum normal load factor (sample interval) | g | - |
| 30 | Mean vertical acceleration (sample interval) | g | $m/s^2$ |
| 31 | Maximum vertical acceleration (sample interval) | g | $m/s^2$ |
| 32 | Average air temperature (ambient, undisturbed air) | °C | K |
| 33 | Atmospheric pressure (air pressure dynamic, start of event) | $lbf/ft^2$ | Pa |
| 34 | Atmospheric pressure (air pressure total, start of event) | hPa | Pa |
| 35 | Atmospheric pressure (air pressure total, start of event) | $lbf/ft^2$ | Pa |
| 36 | Air temperature (total) at start of event | °C | K |
| 37 | Air temperature (total) at start of event (probe 2) | °C | K |
| 38 | Headwind at start of event | knot | m/s |
| 39 | Crosswind at start of event | knot | m/s |
| 40 | Wind direction (true) start of event | degree | rad |
| 41 | Wind speed at start of event | knot | m/s |
| 42 | Air density (total, start of event) | $kg/m^3$ | $kg/m^3$ |
| 43 | EGT average at start of event | °C | K |
| 44 | EGT: Left inboard engine at start of event | °C | K |
| 45 | EGT: Left outboard engine at start of event | °C | K |
| 46 | EGT: Right inboard engine at start of event | °C | K |
| 47 | EGT: Right outboard engine at start of event | °C | K |
| 48 | EPR: Average, percent of maximum at start of event | - | - |
| 49 | Thrust: Percent of maximum at start of event | - | - |
| 50 | Thrust lever angle (left inboard engine, start of event) | degree | rad |
| 51 | Thrust lever angle (left outboard engine, start of event) | degree | rad |

| Sr. No. | Description | Unit in FDR | SI Unit |
|---|---|---|---|
| 52 | Thrust lever angle (right inboard engine, start of event) | degree | rad |
| 53 | Thrust lever angle (right outboard engine, start of event) | degree | rad |
| 54 | Thrust reversers deployed at start of event (true if $< 0.5$) (even if atleast one thrust reverser deployed) | - | - |
| 55 | EMS thrust per engine (average over all engines at start of event) | lbf | N |
| 56 | EMS thrust per engine (enhanced, average over all engines at start of event) | lbf | N |
| 57 | $N_1$: Average (all engines, percent of maximum) at start of event | - | - |
| 58 | $N_1$: Left inboard engine at start of event (% of maximum) | - | - |
| 59 | $N_1$: Left outboard engine at start of event (% of maximum) | - | - |
| 60 | $N_1$: Right inboard engine at start of event (% of maximum) | - | - |
| 61 | $N_1$: Right outboard engine at start of event (% of maximum) | - | - |
| 62 | Average $N_2$ over all engines at start of event (% of maximum) | - | - |
| 63 | Average $N_3$ over all engines at start of event (% of maximum) | - | - |
| 64 | Flap position at start of event | degree | rad |
| 65 | Slat position at start of event | degree | rad |
| 66 | Elevator position at start of event | degree | rad |
| 67 | Horizontal stabilizer position at start of event | degree | rad |
| 68 | Yaw trim position at start of event | degree | rad |
| 69 | Spoiler position average (left) | degree | rad |
| 70 | Spoiler position average (right) | degree | rad |
| 71 | Spoiler position average (left and right) | degree | rad |
| 72 | Pressure altitude at start of event | feet | m |
| 73 | GPS pressure altitude at start of event (best available) | feet | m |
| 74 | Density altitude at start of event | feet | m |
| 75 | Radio height at start of event | feet | m |
| 76 | Height above takeoff (best estimate) at start of event | feet | m |
| 77 | Height above touchdown (best estimate) at start of event | feet | m |
| 78 | Calibrated airspeed (CAS) at start of event | knot | m/s |
| 79 | Airspeed (true) at start of event | knot | m/s |
| 80 | Mach number at start of event | - | - |
| 81 | Speed of sound at start of event | ft/s | m/s |
| 82 | Speed of sound at start of event | knot | m/s |

| Sr. No. | Description | Unit in FDR | SI Unit |
|---|---|---|---|
| 83 | Pitch attitude (Captain's or only) at start of event | degree | rad |
| 84 | Rate of change of pitch rate at start of event | degree/$s^2$ | rad/$s^2$ |
| 85 | Roll attitude (Captain's or only) at start of event | degree | rad |
| 86 | Rate of change of roll rate at start of event | degree/$s^2$ | rad/$s^2$ |
| 87 | Heading (magnetic) at start of event | degree | rad |
| 88 | Yaw/drift angle | degree | rad |
| 89 | Rate of change of yaw rate at start of event | degree/$s^2$ | rad/$s^2$ |
| 90 | Flight path angle (inertial) at start of event | degree | rad |
| 91 | Lateral acceleration (start of event) | g | m/$s^2$ |
| 92 | Longitudinal acceleration (start of event) | g | m/$s^2$ |
| 93 | Normal load factor (start of event) | - | - |
| 94 | Vertical acceleration (start of event) | g | m/$s^2$ |
| 95 | Landing gear down flag (1 = down) | - | - |
| 96 | Average brake temperature at start of event | °C | K |
| 97 | Latitude at start of event (best available) | degree | rad |
| 98 | Longitude at start of event (best available) | degree | rad |
| 99 | Ground track distance from start of takeoff to start of event | nmi | m |
| 100 | GMT at start of event | hr | s |
| 101 | Time from start (first phase of flight) to start of event | s | s |
| 102 | Gross weight at start of event | lb | kg |
| 103 | CG position at start of event (% of mean) | - | - |
| 104 | Drag at start of event (clean configuration) | lbf | N |
| 105 | Lift at start of event | lbf | N |

# Appendix B

# FDR Data: Airports Considered

Table B.1: Airport characteristics. 'Lat.' stands for latitude, 'Lon.' stands for longitude, and 'Elev.' stands for elevation.

| Sr. No. | Airport | IATA Code | Location/City Served | Lat. (°) | Lon. (°) | AMSL Elev. (ft) |
|---|---|---|---|---|---|---|
| 1 | Amsterdam Airport Schiphol | AMS | Amsterdam | 52.31 | 4.77 | -11 |
| 2 | Suvarnabhumi Airport | BKK | Bangkok | 13.7 | 100.7 | 5 |
| 3 | Ibrahim Nasir International Airport | MLE | Hulhulé island (near Malé) | 4.18 | 73.53 | 6 |
| 4 | Bahrain International Airport | BAH | Manama | 26.27 | 50.63 | 6 |
| 5 | Larnaca International Airport | LCA | Larnaca | 34.88 | 33.63 | 7 |
| 6 | Venice Marco Polo Airport | VCE | Venice | 45.5 | 12.34 | 7 |
| 7 | Miami International Airport | MIA | Miami | 25.8 | -80.28 | 8 |
| 8 | Don Mueang International Airport | DMK | Bangkok | 13.91 | 100.6 | 9 |
| 9 | Nice Côte d'Azur Airport | NCE | Nice | 43.66 | 7.21 | 13 |
| 10 | Fiumicino-Leonardo da Vinci International Airport | FCO | Fiumicino (Rome) | 41.8 | 12.25 | 13 |
| 11 | John F. Kennedy International Airport | JFK | New York City | 40.63 | -73.78 | 13 |
| 12 | San Francisco International Airport | SFO | San Francisco Bay Area | 37.62 | -122.38 | 13 |
| 13 | Djerba - Zarzis International Airport | DJE | Djerba | 33.88 | 10.79 | 14 |

| Sr. No. | Airport | IATA Code | Location/City Served | Lat. (°) | Lon. (°) | AMSL Elev. (ft) |
|---|---|---|---|---|---|---|
| 14 | Barcelona El Prat Airport | BCN | Barcelona | 41.3 | 2.07 | 14 |
| 15 | Gregorio Luperón International Airport | POP | Puerto Plata | 19.76 | -70.57 | 15 |
| 16 | Copenhagen Airport, Kastrup | CPH | Copenhagen | 55.61 | 12.64 | 17 |
| 17 | Newark Liberty International Airport | EWR | Newark, NJ | 40.69 | -74.17 | 18 |
| 18 | London City Airport | LCY | London | 51.51 | 0.04 | 19 |
| 19 | General Edward Lawrence Logan International Airport | BOS | Boston | 42.36 | -71.01 | 20 |
| 20 | Thessaloniki International Airport "Macedonia" | SKG | Thessaloniki | 40.52 | 22.98 | 22 |
| 21 | Singapore Changi Airport | SIN | Singapore | 1.36 | 104 | 22 |
| 22 | Palma de Mallorca Airport | PMI | Palma de Mallorca | 39.56 | 2.74 | 24 |
| 23 | Bandaranaike International Airport | CMB | Colombo | 7.18 | 79.88 | 26 |
| 24 | Hazrat Shahjalal International Airport | DAC | Dhaka | 23.85 | 90.4 | 27 |
| 25 | Soekarno-Hatta International Airport | CGK | Tangerang (serves Jakarta area) | -6.12 | 106.6 | 32 |
| 26 | Doha International Airport | DOH | Doha | 25.27 | 51.56 | 35 |
| 27 | Chhatrapati Shivaji International Airport | BOM | Mumbai | 19.09 | 72.87 | 37 |
| 28 | Catania-Fontanarossa Airport | CTA | Catania (island of Sicily) | 37.47 | 15.06 | 39 |
| 29 | Punta Cana International Airport | PUJ | Punta Cana | 18.56 | -68.37 | 40 |
| 30 | King Abdulaziz International Airport | JED | Jeddah | 21.67 | 39.14 | 48 |
| 31 | Muscat International Airport | MCT | Muscat | 23.59 | 58.29 | 49 |
| 32 | Hamburg Airport | HAM | Hamburg | 53.63 | 10 | 53 |
| 33 | Dubai International Airport | DXB | Dubai | 25.26 | 55.35 | 62 |
| 34 | King Fahd International Airport | DMM | Dammam | 26.47 | 49.79 | 72 |
| 35 | Ninoy Aquino International Airport | MNL | Manila | 14.5 | 121 | 75 |
| 36 | Gran Canaria Airport | LPA | Gran Canaria (Canary Islands) | 27.94 | -15.38 | 78 |
| 37 | Julius Nyerere International Airport | DAR | Dar es Salaam | -6.88 | 39.2 | 82 |

| Sr. No. | Airport | IATA Code | Location/City Served | Lat. (°) | Lon. (°) | AMSL Elev. (ft) |
|---|---|---|---|---|---|---|
| 38 | London Heathrow Airport | LHR | London | 51.47 | -0.44 | 83 |
| 39 | Beirut-Rafic Hariri International Airport | BEY | Beirut | 33.83 | 35.5 | 87 |
| 40 | Abu Dhabi International Airport | AUH | Abu Dhabi | 24.42 | 54.66 | 88 |
| 41 | Jinnah International Airport | KHI | Karachi | 24.91 | 67.17 | 100 |
| 42 | Montréal-Pierre Elliott Trudeau International Airport | YUL | Dorval (Montréal) | 45.46 | -73.75 | 118 |
| 43 | Berlin Tegel "Otto Lilienthal" Airport | TXL | Berlin | 52.56 | 13.29 | 122 |
| 44 | Los Angeles International Airport | LAX | Los Angeles | 33.95 | -118.4 | 126 |
| 45 | Heraklion International Airport, "Nikos Kazantzakis" | HER | Heraklion (island of Crete) | 35.34 | 25.19 | 128 |
| 46 | Narita International Airport | NRT | Tokyo | 35.76 | 140.4 | 135 |
| 47 | Sharm el-Sheikh International Airport | SSH | Sharm el-Sheikh | 27.97 | 34.39 | 143 |
| 48 | Düsseldorf Airport | DUS | Düsseldorf | 51.28 | 6.77 | 147 |
| 49 | Antalya Airport | AYT | Antalya | 36.91 | 30.8 | 177 |
| 50 | Hannover-Langenhagen Airport | HAJ | Langenhagen (Hannover) | 52.47 | 9.71 | 183 |
| 51 | Brussels Airport | BRU | Brussels | 50.9 | 4.49 | 184 |
| 52 | Gatwick Airport | LGW | Crawley, West Sussex (serves London) | 51.15 | -0.17 | 203 |
| 53 | Kuwait International Airport | KWI | Kuwait City | 29.23 | 47.97 | 206 |
| 54 | Manchester Airport | MAN | Manchester | 53.36 | -2.28 | 257 |
| 55 | Tozeur-Nefta International Airport | TOE | Tozeur | 33.94 | 8.11 | 287 |
| 56 | Athens International Airport "Eleftherios Venizelos" | ATH | Athens | 37.92 | 23.94 | 308 |
| 57 | Birmingham Airport | BHX | Birmingham | 52.45 | -1.74 | 328 |
| 58 | Warsaw Chopin Airport | WAW | Warsaw | 52.17 | 20.96 | 361 |
| 59 | Frankfurt am Main Airport | FRA | Frankfurt | 50.05 | 8.59 | 364 |
| 60 | Cairo International Airport | CAI | Cairo | 30.11 | 31.41 | 382 |
| 61 | Paris Charles de Gaulle Airport | CDG | Paris | 49 | 2.56 | 392 |

| Sr. No. | Airport | IATA Code | Location/City Served | Lat. (°) | Lon. (°) | AMSL Elev. (ft) |
|---|---|---|---|---|---|---|
| 62 | Kos Island International Airport, "Hippocrates" | KGS | Kos island | 36.8 | 27.09 | 409 |
| 63 | Budapest Ferenc Liszt International Airport | BUD | Budapest | 47.42 | 19.25 | 495 |
| 64 | London Luton Airport | LTN | Luton, Bedfordshire (serves London) | 51.88 | -0.37 | 526 |
| 65 | Toronto Pearson International Airport | YYZ | Toronto | 43.69 | -73.64 | 569 |
| 66 | Vienna International Airport | VIE | Vienna | 48.11 | 16.57 | 600 |
| 67 | Mohammed V International Airport | CMN | Casablanca | 33.38 | -7.59 | 656 |
| 68 | Chicago O'Hare International Airport | ORD | Chicago | 41.97 | -87.89 | 668 |
| 69 | Allama Iqbal International Airport | LHE | Lahore | 31.52 | 74.41 | 698 |
| 70 | Milano Malpensa Airport | MXP | Milan | 45.61 | 8.73 | 768 |
| 71 | Indira Gandhi International Airport | DEL | New Delhi | 28.56 | 77.09 | 777 |
| 72 | EuroAirport Basel-Mulhouse-Freiburg | BSL/ MLH/ EAP | Basel/ Mulhouse/ Freiburg | 47.61 | 7.52 | 885 |
| 73 | Bacha Khan International Airport | PEW | Peshawar | 33.99 | 71.52 | 1158 |
| 74 | Luxembourg Findel Airport | LUX | Luxembourg City | 49.63 | 6.23 | 1234 |
| 75 | Prague Václav Havel Airport | PRG | Prague | 50.11 | 14.27 | 1247 |
| 76 | Khartoum International Airport | KRT | Khartoum | 15.59 | 32.55 | 1265 |
| 77 | Stuttgart Airport | STR | Stuttgart | 48.69 | 9.24 | 1276 |
| 78 | Geneva International Airport | GVA | Geneva | 46.24 | 6.11 | 1411 |
| 79 | Zurich Airport | ZRH | Zurich | 47.46 | 8.56 | 1416 |
| 80 | Munich Airport | MUC | Munich | 48.35 | 11.78 | 1487 |
| 81 | Benazir Bhutto International Airport | ISB | Islamabad, Rawalpindi | 33.61 | 73.11 | 1688 |
| 82 | Madrid-Barajas Airport | MAD | Madrid | 40.46 | -3.55 | 2000 |
| 83 | Damascus International Airport | DAM | Damascus | 33.41 | 36.52 | 2020 |
| 84 | King Khalid International Airport | RUH | Riyadh | 24.96 | 46.7 | 2049 |
| 85 | Prince Mohammad Bin Abdulaziz International Airport | MED | Medina | 24.55 | 39.7 | 2151 |

| Sr. No. | Airport | IATA Code | Location/City Served | Lat. (°) | Lon. (°) | AMSL Elev. (ft) |
|---|---|---|---|---|---|---|
| 86 | Queen Alia International Airport | AMM | Amman | 31.73 | 36 | 2395 |
| 87 | São Paulo/Guarulhos-Governador André Franco Montoro International Airport | GRU | São Paulo | -23.43 | -46.48 | 2459 |
| 88 | Jomo Kenyatta International Airport | NBO | Nairobi | -1.32 | 36.93 | 5327 |
| 89 | O.R. Tambo International Airport | JNB | Kempton Park (serves Johannes-burg, Preto-ria) | -26.13 | 28.23 | 5558 |

THIS PAGE IS INTENTIONALLY LEFT BLANK.

# Bibliography


[1] Gabriele Enea, Jesper Bronsvoort, Hartmut Fricke, Christian Seiß, Judith Rosenow, Almira Ramadani, and Mike Paglione. Fuel burn estimation modeling for ATM benchmark applications perspectives from an international collaboration. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2017.

[2] International Air Transport Association (IATA). Fact sheet: Economic and social benefits of air transport. http://www.iata.org/pressroom/facts_figures/fact_sheets/Documents/fact-sheet-economic-and-social-benefits-of-air-transport.pdf, 2017. [Online; accessed June 2017].

[3] International Air Transport Association (IATA). Fact sheet: Industry statistics. http://www.iata.org/pressroom/facts_figures/fact_sheets/Documents/fact-sheet-industry-facts.pdf, 2017. [Online; accessed June 2017].

[4] FAA Office of Environment and Energy. Aviation emissions, impact, and mitigation: A primer, 2015.

[5] Bethan Owen, David S. Lee, and Ling Lim. Flying into the future: Aviation emissions scenarios to 2050. *Environmental Science and Technology*, 44(7):2255–2260, March 2010.

[6] Federal Aviation Administration. Advisory circular: Flight Operational Quality Assurance. http://www.airweb.faa.gov/Regulatory_and_Guidance_Library/rgAdvisoryCircular.nsf/0/40c02fc39c1577b686256e8a005afb0a/\protect\T1\textdollarFILE/AC120-82.pdf, April 2004. [Online; accessed July 2017].

[7] Shelley E. Lowe, Elaine M. Pfleiderer, and Thomas R. Chidester. Perceptions and efficacy of Flight Operational Quality Assurance (FOQA) programs among small-scale operators. http://libraryonline.erau.edu/online-full-text/faa-aviation-medicine-reports/AM12-01.pdf, January 2012. Report No. DOT/FAA/AM-12/1, [Online; accessed July 2017].

[8] International Air Transport Association (IATA). Another strong year for airline profits in 2017. http://www.iata.org/pressroom/pr/Pages/2016-12-08-01.aspx, 2016. [Online; accessed July 2016].

[9] Bela P. Collins. Estimation of aircraft fuel consumption. *AIAA Journal of Aircraft*, 19(11):969–975, November 1982.

[10] M. Ahearn et al. *Aviation Environmental Design Tool (AEDT) Technical Manual, Version 2b, Service Pack 3*. U.S. Department of Transportation John A. Volpe National Transportation Systems Center, 2016. Report No. DOT-VNTSC-FAA-16-11.

[11] Joint Planning and Development Office (JPDO). Next generation air transportation system integrated plan, December 2004.

[12] SESAR Joint Undertaking. SESAR website. https://www.sesarju.eu/. [Online; accessed July 2017].

[13] Australian Government, Department of Infrastructure and Regional Development. Australia's air traffic management plan-draft, 2016.

[14] Federal Aviation Administration. Aviation Environmental Design Tool (AEDT) website. https://aedt.faa.gov/. [Online; accessed June 2017].

[15] Robert D. Oaks, Hollis F. Ryan, and Mike Paglione. Prototype implementation and concept validation of a 4-d trajectory fuel burn model application. In *AIAA Guidance, Navigation, and Control Conference*, August 2010.

[16] Michael Kaiser, Michael Schultz, and Hartmut Fricke. Enhanced jet performance model for high precision 4d flight path prediction. In *ATACCS-2011*, pages 33–40, May 2011.

[17] Jesper Bronsvoort, Paul Zissermann, Steven Barry, and Greg McDonald. A framework for assessing and managing the impact of ANSP actions on flight efficiency. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2015.

[18] Gano B. Chatterji. Fuel burn estimation using real track data. In *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, September 2011.

[19] John E. Robinson III and Maryam Kamgarpour. Benefits of continuous descent operations in high-density terminal airspace under scheduling constraints. In *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, September 2010.

[20] Luke Jensen, Henry Tran, and R. John Hansman. Cruise fuel reduction potential from altitude and speed optimization in global airline operations. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2015.

[21] EUROCONTROL. Base of Aircraft Data (BADA) website. http://www.eurocontrol.int/services/bada. [Online; accessed June 2017].

[22] A. Nuic. *User Manual for the Base of Aircraft Data (BADA) Revision 3.13*. EUROCONTROL Experimental Centre, 2015. EEC Technical/Scientific Report No. 15/04/02-43.

[23] EUROCONTROL. Base of Aircraft Data (BADA) fact-sheet. `http://www.eurocontrol.int/publications/base-aircraft-data-bada-factsheet`. [Online; accessed June 2017].

[24] Eduardo Gallo, Francisco A. Navarro, Angela Nuic, and Mihai Iagaru. Advanced aircraft performance modelling for ATM: BADA 4.0 results. In *25th Digital Avionics Systems Conference*, October 2006.

[25] EUROCONTROL Experimental Centre. *Base of Aircraft Data (BADA) Aircraft Performance Modelling Report*, 2009. EEC Technical/Scientific Report No. 2009-009.

[26] John Heimlich. Coping with sky-high jet fuel prices, 2008.

[27] Boeing Commercial Airplanes – Flight Operations Engineering. Fuel conservation, 2004.

[28] David A. Senzig, Gregg G. Fleming, and Ralph J. Iovinelli. Modeling of terminal-area airplane fuel consumption. *AIAA Journal of Aircraft*, 46(4):1089–1093, July-August 2009.

[29] Hartmut Fricke, Christian Seiß, and Robert Herrmann. Fuel and energy benchmark analysis of continuous descent operations. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2015.

[30] U.S. Department of Transportation. The Next Generation air transportation system (NextGen). `www.transportation.gov/mission/sustainability/next-generation-air-transportation-system-nextgen`, 2014. [Online; accessed August 2017].

[31] Gennady G. Kulikov and Haydn A. Thompson, editors. *Dynamic Modelling of Gas Turbines: Identification, Simulation, Condition Monitoring, and Optimal Control*. Springer-Verlag, London, 2004.

[32] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2009.

[33] Hak-Tae Lee and Gano B. Chatterji. Closed-form takeoff weight estimation model for air transportation simulation. In *AIAA Aviation Technology, Integration, and Operations (ATIO) Conference*, September 2010.

[34] EUROCONTROL Experimental Centre. Aircraft Noise and Performance (ANP) Database. [Online; accessed January 2017].

[35] Yashovardhan S. Chati and Hamsa Balakrishnan. Data-driven modeling of aircraft engine fuel burn in climb out and approach. *Transportation Research Record: Journal of the Transportation Research Board*, 2018 (to appear).

[36] Yashovardhan S. Chati and Hamsa Balakrishnan. A Gaussian Process Regression approach to model aircraft engine fuel flow rate. In *ACM/IEEE International Conference on Cyber-Physical Systems*, April 2017. DOI: 10.1145/3055004.3055025.

[37] Yashovardhan S. Chati and Hamsa Balakrishnan. Statistical modeling of aircraft takeoff weight. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2017.

[38] GasTurb 12 software. `http://gasturb.de/`, 2015. [Online; accessed October 2015].

[39] Nicholas A. Cumpsty. *Jet Propulsion: A Simple Guide to the Aerodynamics and Thermodynamic Design and Performance of Jet Engines*. Cambridge University Press, Cambridge, UK, 2nd edition, 2003.

[40] National Science Foundation. Cyber-Physical Systems. `www.nsf.gov/funding/pgm_summ.jsp?pims_id=503286&org=CISE`, 2017. [Online; accessed August 2017].

[41] National Institute of Standards and Technology. Cyber-Physical Systems. `www.nist.gov/el/cyber-physical-systems`, 2017. [Online; accessed August 2017].

[42] Liang He, Eugene Kim, Kang G. Shin, Guozhu Meng, and Tian He. Battery state-of-health estimation for mobile devices. In *ACM/IEEE International Conference on Cyber-Physical Systems*, April 2017.

[43] Dimitri P. Solomatine. *Data-Driven Modeling and Computational Intelligence Methods in Hydrology*. John Wiley & Sons, Ltd., 2006.

[44] Madhur Behl, Achin Jain, and Rahul Mangharam. Data-driven modeling, control and tools for cyber-physical energy systems. In *ACM/IEEE International Conference on Cyber-Physical Systems*, April 2016.

[45] Dali Wei. *Data-Driven Modeling and Transportation Data Analytics*. PhD thesis, Texas Tech University, May 2014.

[46] Antonio A. Trani, F. C. Wing-Ho, Glen Schilling, Hojong Baik, and Anand Seshadri. A neural network model to estimate aircraft fuel consumption. In *AIAA A neural network model to estimate aircraft fuel consumption Aviation Technology, Integration and Operations (ATIO) Forum*, September 2004.

[47] David A. Senzig, Gregg G. Fleming, and Ralph J. Iovinelli. Fuel consumption modeling in support of ATM environmental decision-making. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2009.

[48] International Civil Aviation Organization. ICAO Aircraft Engine Emissions Databank. `https://easa.europa.eu/document-library/icao-aircraft-engine-emissions-databank`, 2014. [Online; accessed 12 February 2014].

[49] Yashovardhan S. Chati and Hamsa Balakrishnan. Analysis of aircraft fuel burn and emissions in the landing and take off cycle using operational data. In *International Conference on Research in Air Transportation*, May 2014.

[50] Judith Patterson, George J. Noel, David A. Senzig, Christopher J. Roof, and Gregg G. Fleming. Analysis of departure and arrival profiles using real-time aircraft data. *AIAA Journal of Aircraft*, 46(4):1094–1103, July-August 2009.

[51] Scott C. Herndon, Todd Rogers, Edward J. Dunlea, John T. Jayne, Richard Miake-Lye, and Berk Knighton. Hydrocarbon emissions from in-use commercial aircraft during airport operations. *Environmental Science and Technology*, 40(14):4406–4413, July 2006.

[52] Scott C. Herndon, Ezra C. Wood, Megan J. Northway, Richard Miake-Lye, Lee Thornhill, Andreas Beyersdorf, Bruce E. Anderson, Renee Dowlin, William Dodds, and W. Berk Knighton. Aircraft hydrocarbon emissions at Oakland International Airport. *Environmental Science and Technology*, 43(6):1730–1736, February 2009.

[53] Megan S. Ryerson, Mark Hansen, and James Bonn. Validating aircraft performance models with airline data. In *International Conference on Research in Air Transportation*, May 2012.

[54] Harshad Khadilkar and Hamsa Balakrishnan. Estimation of aircraft taxi fuel burn using flight data recorder archives. *Transportation Research Part D: Transport and Environment*, 17(7):532–537, 2012.

[55] Brian Kim, Gregg Fleming, Sathya Balasubramanian, Andrew Malwitz, Joosung Lee, Joseph Ruggiero, Ian Waitz, Kelly Klima, Virginia Stouffer, Dou Long, Peter Kostiuk, Maryalice Locke, Curtis Holsclaw, Angel Morales, Edward McQueen, and Warren Gillette. *System for assessing Aviation's Global Emissions (SAGE), Version 1.5, Technical Manual.* John A. Volpe National Transportation Systems Center, Massachusetts Institute of Technology, and Logistics Management Institute, September 2005. Report No. DOT-VNTSC-FAA-05-14.

[56] Tim Yoder. Development of aircraft fuel burn modeling techniques with applications to global emissions modeling and assessment of the benefits of reduced vertical separation minimums. Master's thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA, 2007.

[57] Alan J. Stolzer. Fuel consumption modeling of a transport category aircraft: A Flight Operations Quality Assurance (FOQA) analysis. *Journal of Air Transportation*, 8(2):3–18, 2003.

[58] E.G. Tulapurkara. National Programme on Technology Enhanced Learning (NPTEL) India: Lectures on airplane design (aerodynamic): Weight estimation, 2013. [Online; accessed January 2017].

[59] Daniel P. Raymer. *Aircraft Design: A Conceptual Approach.* American Institute of Aeronautics and Astronautics (AIAA), Reston, Virginia, 5th edition, 2012.

[60] Jan Roskam. *Airplane Design Part V: Component Weight Estimation.* Roskam Aviation and Engineering Corp., Ottawa, Kansas, 1985.

[61] Lance Sherry and Saba Neyshabouri. Estimating takeoff thrust from surveillance track data. In *Transportation Research Board Annual Meeting*, January 2014.

[62] U.S. Department of Transportation. Air Carrier Statistics Database, 2016.

[63] Richard Alligier, David Gianazza, and Nicolas Durand. Energy rate prediction using an equivalent thrust setting profile. In *International Conference on Research in Air Transportation*, May 2012.

[64] Richard Alligier, David Gianazza, and Nicolas Durand. Ground-based estimation of the aircraft mass, adaptive vs. least squares method. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2013.

[65] Richard Alligier, David Gianazza, and Nicolas Durand. Learning the aircraft mass and thrust to improve the ground-based trajectory prediction of climbing flights. *Transportation Research Part C: Emerging Technologies*, 36:45–60, 2013.

[66] Richard Alligier, David Gianazza, Mohammad Ghasemi Hamed, and Nicolas Durand. Comparison of two ground-based mass estimation methods on real data. In *International Conference on Research in Air Transportation*, May 2014.

[67] Richard Alligier, David Gianazza, and Nicolas Durand. Machine learning and mass estimation methods for ground-based aircraft climb prediction. *IEEE Transactions on Intelligent Transportation Systems*, 16(6):3138–3149, December 2015.

[68] Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. Modeling and inferring aircraft takeoff mass from runway ADS-B data. In *International Conference on Research in Air Transportation*, June 2016.

[69] Junzi Sun, Joost Ellerbroek, and Jacco Hoekstra. Bayesian inference of aircraft initial mass. In *USA/Europe Air Traffic Management Research and Development (ATM) Seminar*, June 2017.

[70] Paul Jackson, editor. *Jane's All the World's Aircraft: Development and Production.* Jane's by IHS Markit, 2017-18.

[71] Jamie Hunter, editor. *Jane's All the World's Aircraft: In Service.* Jane's by IHS Markit, 2017-18.

[72] Paul Jackson, editor. *Jane's All the World's Aircraft.* International Thomson Publishing Company, 1995-96.

[73] Bucher and Company. JP Airline Fleets International (2006/2007), 2007.

[74] Bucher and Company. JP Airline Fleets International (2015/2016), 2016.

[75] Sensis Corporation, East Syracuse, NY. *ASDE-X brochure*, 2008.

[76] Michael H. Kutner, Christopher J. Nachtsheim, John Neter, and William Li. *Applied Linear Statistical Models*. McGraw Hill Education (India) Private Limited, New Delhi, 5$^{th}$ edition, 2013.

[77] John K. Kruschke. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, San Diego, California, 2$^{nd}$ edition, 2015.

[78] The Mathworks, Inc., MATLAB® R2012a documentation, 2012. Software package.

[79] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification and Regression Trees*. CRC Press, New York, 1993.

[80] Yashovardhan S. Chati and Hamsa Balakrishnan. Statistical modeling of aircraft engine fuel flow rate. In *Congress of the International Council of the Aeronautical Sciences*, September 2016.

[81] Hiroyuki Mori and Akira Takahashi. A data mining method for selecting input variables for forecasting model of global solar radiation. In *Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES*, pages 1–6, May 2012.

[82] Dillon Matthew Carty. An analysis of boosted regression trees to predict the strength properties of wood composites. Master's thesis, University of Tennessee, USA, 2011.

[83] Andreas Hecker and Thomas Kürner. Application of classification and regression trees for paging traffic prediction in LAC planning. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 874–878, April 2007.

[84] J. Richard Stewart. Applications of classification and regression tree methods in roadway safety studies. *Transportation Research Record: Journal of the Transportation Research Board*, 1542:1–5, 1996.

[85] Wlamir O.L. Vianna, João P.P. Gomes, Roberto K.H. Galvão, Takashi Yoneyama, and Jackson P. Matsuura. Health monitoring of an auxiliary power unit using a classification tree. In *Annual Conference of the Prognostics and Health Management Society*, 2011.

[86] Wei Cheng, Xiuying Zhang, Ke Wang, and Xuelong Dai. Integrating classification and regression tree (CART) with GIS for assessment of heavy metals pollution. *Environmental Monitoring and Assessment*, 158(1-4):419–431, November 2009.

[87] J. Elith, J.R. Leathwick, and T. Hastie. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4):802–813, July 2008.

[88] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Springer, 1993.

[89] Robert Dürichen, Marco A.F. Pimentel, Lei Clifton, Achim Schweikard, and David A. Clifton. Multi-task Gaussian Process models for biomedical applications. In *IEEE-EMBS International Conference on Biomedical and Health Informatics*, June 2014.

[90] Lei Clifton, David A. Clifton, Marco A.F. Pimentel, Peter J. Watkinson, and Lionel Tarassenko. Gaussian Process Regression in vital-sign early warning systems. In *Annual International Conference of the IEEE EMBS*, August 2012.

[91] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko. Probabilistic estimation of respiratory rate using Gaussian Processes. In *Annual International Conference of the IEEE EMBS*, July 2013.

[92] Carolynne Hultquist, Gang Chen, and Kaiguang Zhao. A comparison of Gaussian Process Regression, Random Forests, and Support Vector Regression for burn severity assessment in diseased forests. *Remote Sensing Letters*, 5(8):723–732, 2014.

[93] Luca Pasolli, Farid Melgani, and Enrico Blanzieri. Gaussian Process Regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 7(3):464–468, July 2010.

[94] John C. Platt, Christopher J.C. Burges, Steven Swenson, Christopher Weare, and Alice Zheng. Learning a Gaussian Process prior for automatically generating music playlists. In *Advances in Neural Information Processing Systems 14*, pages 1425–1432, 2002.

[95] Kian Ming A. Chai, Christopher K.I. Williams, Stefan Klanke, and Sethu Vijayakumar. Multi-task Gaussian Process learning of robot inverse dynamics. In *Advances in Neural Information Processing Systems 21*, pages 265–272, 2009.

[96] Anton Schwaighofer, Marian Grigoraş, Volker Tresp, and Clemens Hoffmann. GPPS: A Gaussian Process positioning system for cellular networks. In *Advances in Neural Information Processing Systems 16*, pages 579–586, 2004.

[97] C.A.L. Bailer-Jones, T.J. Sabin, D.J.C. MacKay, and P.J. Withers. Prediction of deformed and annealed microstructures using Bayesian neural networks and Gaussian Processes. In *Proceedings of the Australasia-Pacific Forum on Intelligent Processing and Manufacturing of Materials*, 1997.

[98] Carl Edward Rasmussen and Christopher K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[99] Jarno Vanhatalo, Jaakko Riihimäki, Jouni Hartikainen, Pasi Jylänki, Ville Tolvanen, and Aki Vehtari. Bayesian modeling with Gaussian Processes using the GPstuff toolbox. In *arXiv:1206.5754 [stat.ML]*, 2015.

[100] Joaquin Quiñonero-Candela and Carl Edward Rasmussen. A unifying view of sparse approximate Gaussian Process Regression. *Journal of Machine Learning Research*, 6:1939–1959, 2005.

[101] Jarno Vanhatalo, Jaakko Riihimäki, Jouni Hartikainen, Pasi Jylänki, Ville Tolvanen, and Aki Vehtari. GPstuff: Bayesian modeling with Gaussian Processes. *Journal of Machine Learning Research*, 14:1175–1179, 2013.

[102] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, December 1945.

[103] John David Anderson, Jr. *Introduction to Flight*. McGraw-Hill Higher Education, 5th international edition, 2004.

[104] National Oceanic and Atmospheric Administration. METARs. `aviationweather.gov/metar?gis=off`, 2017. [Online; accessed August 2017].

[105] National Oceanic and Atmospheric Administration. Rapid Refresh. `www.ncdc.noaa.gov/data-access/model-data/model-datasets/rapid-refresh-rap`, 2017. [Online; accessed August 2017].

[106] National Oceanic and Atmospheric Administration. World Area Forecast System. `www.aviationweather.gov/wifs/`, 2017. [Online; accessed August 2017].

[107] Agathe Girard, Carl Edward Rasmussen, Joaquin Quiñonero-Candela, and Roderick Murray-Smith. Gaussian Process priors with uncertain inputs-application to multiple-step ahead time series forecasting. In *Advances in Neural Information Processing Systems 15*, pages 545–552, 2003.

[108] Sandeep Badrinath. ASDE-X smoothing algorithms. Personal communication, 2017.

[109] Mark Briers, Arnaud Doucet, and Simon Maskell. Smoothing algorithms for state–space models. *Annals of the Institute of Statistical Mathematics*, 62(1):61–89, June 2009.

[110] Rémy Lopez and Patrick Danès. Low-complexity IMM smoothing for jump Markov nonlinear systems. *IEEE Transactions on Aerospace and Electronic Systems*, 53(3):1261–1272, June 2017.

[111] U.S. Department of Transportation Federal Aviation Administration. *Aviation Environmental Design Tool (AEDT) 2a Uncertainty Quantification Report*, 2014. Report No. DOT/FAA/AEE/2013-03.

[112] National Oceanic and Atmospheric Administration. Global Forecast System (GFS). `https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs`, 2017. [Online; accessed August 2017].